



Application of Deep Learning to Prostate Multi-parametric MRI : Segmentation, Lesion Detection and Classification

Mémoire présenté en vue de l'obtention du diplôme
d'Ingénieur Civil en informatique à finalité spécialisée

Naim Sassine

Promoteur
Professeur Olivier Debeir

Service
LISA

Année académique
2020 - 2021

Acknowledgements

As I am completing this work of research, I would like to address a special thank you to Professor Olivier Debeir which supported me and guided me throughout this entire work. In these times of crisis where students are struggling to find motivation and purpose, Professor Debeir assisted me on a moral plan and on a work ethic plan. With the help he provided me, I was able to face many different challenges and complete this work of research. I would also like to thank the entire LISA team for providing me with technical help, assistance and material.

Abstract

Le cancer de la prostate est l'un des cancers les plus répandus dans le monde. Des techniques invasives sont généralement utilisées pour diagnostiquer le cancer de la prostate. Les méthodes non invasives ont l'avantage de causer très peu de dommages aux patients et de produire des résultats plus rapides. L'imagerie par résonance magnétique multi-paramétrique (IRM-mp) est une technique non invasive utilisée pour détecter, classer et quantifier les lésions de la prostate. Les images produites sont généralement étudiées par les radiologues afin de diagnostiquer le cancer de la prostate. Ce diagnostic souffre d'une détection ou d'une interprétation erronée en raison de facteurs tels que les limites de l'observateur, la qualité de l'image, la complexité des cas cliniques et la variabilité de l'apparence des lésions. Le Deep Learning (DL), un sous-ensemble de l'intelligence artificielle, est une technologie qui est actuellement utilisée pour découvrir et étudier les motifs complexes qui se cachent derrière les caractéristiques des images. Le DL a le potentiel de fournir une meilleure standardisation et cohérence dans l'identification des lésions de la prostate et d'améliorer le diagnostic du cancer de la prostate en étudiant les scans IRM-mp. Le but de cette recherche est de mettre en place un logiciel basé sur du Deep Learning qui est construit, entraîné et testé pour produire des informations qui pourraient potentiellement aider les médecins à améliorer leurs diagnostics du cancer de la prostate. Les trois tâches principales que ce logiciel réalise sont la segmentation de la prostate, la détection des lésions et la classification des lésions en lésions significatives ou non significatives. Deux réseaux DL ont été conçus pour accomplir ces trois tâches, chacun obtenant un Dice Similarity Coefficient (DSC) de 87% et 61% respectivement. Ces résultats montrent que ce logiciel basé sur deux modèles DL fournit des prédictions précises et des informations qui pourraient être utilisées par les médecins pour améliorer leur diagnostic du cancer de la prostate. L'intégrité du code source développé pour ce travail de recherche peut être trouvée ici : <https://github.com/naimsassine/MasterThesisInfo2021>

Mots-clés : Deep Learning, Cancer de la Prostate, IRM-mp, Segmentation, Classification, Lesions Prostatiques

Abstract

Prostate carcinoma or prostate cancer is one of the most prevalent cancers worldwide. Invasive techniques are usually used to diagnose prostate cancer. Non-invasive methods have the advantage of causing very little damage to the patients and producing faster results. Multi-parametric magnetic resonance imaging (mp-MRI) is a non-invasive technique used for detecting, classifying, and quantifying prostate lesions. The produced images are generally studied by radiologists in order to diagnose prostate cancer. This diagnosis suffers from erroneous detection or interpretation due to factors like the observer's limitations, the image quality, the complexity of the clinical cases and variability in lesions appearance. Deep Learning (DL), a subset of Artificial Intelligence is a technology that is currently used to uncover and study the patterns behind features of images. DL has a potential of providing better standardization and consistency in identifying prostate lesions and improving prostate cancer diagnosis by studying mp-MRI scans. In this research, a computer aided software based on deep learning is built, trained and tested to produce insights that could potentially help doctors improve their diagnoses of prostate cancer. The three main tasks that this software achieves are prostate segmentation, lesion detection and lesion classification into significant or non-significant lesions. Two DL networks were conceived to accomplish these three tasks, each achieving a Dice Similarity Coefficient (DSC) of 87% and 61% respectively. These results show that this built software based on two DL models provides accurate predictions and insights that could be used by doctors to improve their diagnosis of prostate cancer. The integrity of the source code developed during this work of research can be found here : <https://github.com/naimsassine/MasterThesisInfo2021>

Keywords : Deep Learning, Prostate Cancer, mp-MRI, Segmentation, Classification, Prostatic Lesions

Contents

1	Introduction	2
1.1	Context	2
1.2	Related Work	6
1.2.1	Prostate Segmentation	6
1.2.2	Lesion Detection	8
1.2.3	Lesion Classification	9
1.2.4	Grad-CAM	10
1.3	Research Objectives	12
2	Materials and Methods	14
2.1	MRI Imaging Techniques	14
2.1.1	T2-W-MRI	15
2.1.2	DCE-MRI	15
2.1.3	DW-MRI	16
2.1.4	PDW	16
2.2	MRI Datasets	18
2.2.1	Prostate Segmentation	18
2.2.2	Lesion Detection and Classification	19
2.3	Deep Learning Pipeline	21
2.3.1	U-Net Architecture	21
2.3.2	Prostate Segmentation	23
2.3.3	Lesion Detection and Classification	27
2.3.4	Entire Pipeline	36
3	Experiments and Results	39
3.1	Prostate Segmentation	41
3.2	Lesion Detection and Classification	46
3.3	Entire Pipeline	52
4	Conclusion	54

Chapter 1

Introduction

1.1 Context

The prostate is a walnut-sized male organ that acts as both an accessory gland of the male reproductive system and a muscular switch between urination and ejaculation. As an accessory gland, it produces and contains essential fluid that is part of the semen. Anatomically, the prostate is localised below the bladder, with the urethra passing through it. It usually is divided in different zones or lobes, precisely the central zone, the peripheral zone and the transitional zone [1].

Prostatic carcinoma, or better known as Prostate Cancer (PCa) is a type of cancer that develops in the prostate gland. Just like other types of cancers, this particular disease happens when abnormal cells divide uncontrollably and destroy healthy tissue. PCa induces growth to the prostate gland, but it should not be confused with Benign Prostatic Hyperplasia (BPH). In fact, growths in this male gland can be benign (classified as not cancerous) or malignant (classified as cancerous) [1].

- Benign growths, which could be induced by BPH, are generally not a threat to life. They don't spread to other parts of the body and they don't invade the tissues around them.
- Malignant growths, which are synonyms to prostate cancer, are of a much bigger threat to life than Benign growths. They can spread to nearby organs as well as other parts of the body.

Studies made in 2020 have shown that [2] on a worldwide scale, prostate cancer is the second most frequently diagnosed cancer for men, accounting for a total of 14.1% of male cancer cases. With 1,414,259 estimated cases worldwide and 375,304 deaths in 2020, PCa is known to be a commonly diagnosed and dangerous cancer among men, if not detected and treated early on.

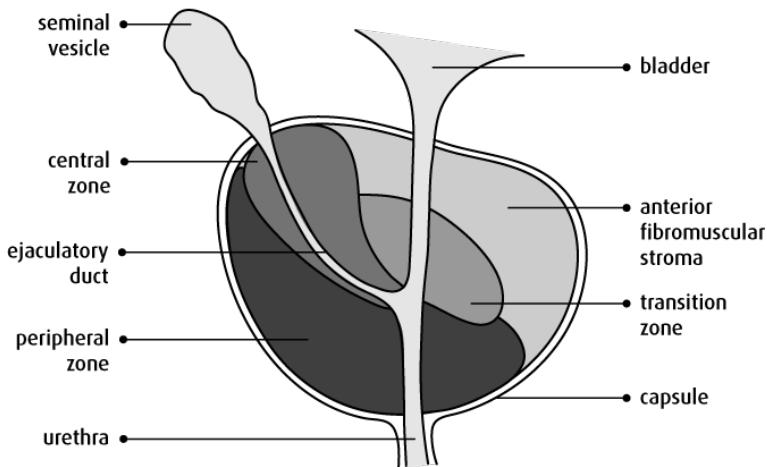


Figure 1.1: Zones of the Prostate [3]

In its starting stages, prostate cancer often shows no symptoms. They usually appear a few months later and are quite similar to symptoms of an enlarged prostate or BPH. These symptoms range from urinary problems and painful ejaculations to bone and lower back pain [1].

Currently, no direct causes are known. According to autopsy studies, one in every three men over the age of 50 had cancer cells in their prostate [1]. Even though we don't know of a direct cause for PCa, we know that many risks are associated with the diagnosis of the disease : Age, Ethnicity, Genetics, Smoking, Diet [1] ...

In terms of PCa diagnosis, the different techniques can be grouped into two categories. The first is screening, which involves being tested for a disease even if you have no symptoms, such as at a routine appointment. The second is biopsy. It is a forward step into the diagnosis of PCa and is looked into after screening specially when the doctor is suspicious of a cancer [1].

1. For PCa, screening methods include prostate specific antigen (PSA) blood test and digital rectal examination (DRE). Despite the fact that both tests are used to detect cancer early, abnormal results from either test could be due to BPH or an infection rather than malignancy. The PSA test measures the amount of prostate-specific antigen in the blood, which is a protein characteristic of prostate cancer and prostate diseases. The DRE helps the doctor checkup on the prostate. The prostate sits in front of the rectum, which enables the doctors to feel the prostate by inserting a gloved lubricated finger into the rectum. The doctor may be able to tell whether there are any hard, lumpy, or atypical spots on the prostate by feeling it. These could be symptoms of anomalies. These screening methods are essential to be able to detect PCa early on, but as we said previously, other types of prostate anomalies could lead to similar screening results
2. A more advanced step in the diagnosis would be the biopsy. This invasive procedure (type of minor surgery) consists in going into the body of the patient and extracting

very small pieces of prostate tissue. These samples are later analysed carefully by a pathologist in a laboratory under a microscope. This procedure is the most accurate way to determine if your prostate contains cancerous cells. The PSA values and DRE results are usually used to determine whether or not a biopsy should be performed.

To define the progress of a patient's prostate cancer a common metric that is used is grading. Grading is done based on the biopsy samples collected and analysed in the laboratories. When a doctor studies these tissues, he deduces a grading describing how fast the cancerous cells (if detected) are likely to evolve and spread. The goal of this grading is to label these cancerous regions by their aggressiveness [1].

The most used grading system is called the Gleason Grading System. The grade is assigned by studying the arrangement of the cells in the prostate. The grade is given on a scale from 1 to 5, 1 and 2 meaning no cancerous tissue was detected, 3 suggesting a slow growing tumour and a grade of 5 indicating a highly aggressive, high-risk form of prostate cancer. The two most prevalent grades observed in biopsy samples are then combined to get a final score (score of grades $3+3 = 6$ means that its a slow growing cancer, $5+5 = 10$ means that the cancer is very aggressive). This given score will help doctors evaluate how dangerous the cancer is and will certainly impact the treatment decisions [1].

In terms of treatment, the doctors usually study the aggressiveness of the cancer before suggesting a specific treatment track. They do that to be able to predict at what rate is the prostate cancer growing. A lot of patients suffering from PCa have very slow growing cancers and don't necessarily need any treatment thanks to this very low growth. If this is not the case, the treatment strategy will be based on a variety of criteria and will most likely include surveillance localized therapy (such as surgery, radiation therapy) or systemic therapy (hormonal therapy, Chemotherapy) [1].

As we have seen above, one of the best way to deal with PCa is to detect it in its early stages and observe its progression over time. Screening method discussed previously could be used but as we have seen abnormal screening results don't necessarily lead to prostate cancer. The only way to make sure that cancerous tissues are present in the prostate is by an invasive technique which is the biopsy. The lack of non-invasive techniques that can distinguish between aggressive and non-aggressive cancer types is a key barrier in prostate cancer care. As we know, technology has been shaping its way onto helping the medical field in many different ways, one of them being medical imaging [4]. In fact, a part from being non-invasive techniques, the progress in this field has enabled medical imaging to become essential and beneficial procedures that allow doctors to diagnose diseases and injuries without being intrusive. These techniques include x-rays, computed tomography (CT) scans, and magnetic resonance imaging (MRI).

Prostate MRI has become a more common procedure in everyday radiology procedures, and it is normally performed as a multi-parametric (mp-MRI) or bi-parametric (bi-MRI) MRI of the prostate. The innovations around prostate MRI have allowed doctors and more specially radiologists to detect, assess and analyse PCa based on these images. In fact, a reporting scheme has been developed internationally to evaluate sus-

pected prostate cancer by looking at mp-MRI. This reporting scheme called PI-RADS (Prostate Imaging-Reporting and Data System) [5] was created by the American College of Radiology, AdMeTech Foundation and ESUR as a joint effort with a precise goal : develop a scoring system to aid in the early detection of clinically relevant prostate cancer and the reduction of unnecessary biopsies and care for benign and sub-clinical diseases [5].

However, erroneous abnormality detection or interpretation continues to haunt doctors' and radiologists' diagnosis throughout the clinical process. In fact, these errors are driven by many factors [4]:

- Observer limitations : fatigue, distraction, constrained human visual perception
- Lack of clinical experience
- Image quality
- Complexity of the clinical cases and variability in lesions appearance
- Imbalanced data : radiologists see more healthy cases than malignant
- Overlapping structures within the prostate MRIs
- Large amount of data (time consuming and potentially driving errors)

The revolutionary technologies developed in the field of computer vision and artificial intelligence (AI) have been key to the design of computer aided detection and diagnosis systems (CAD). The purpose of these systems is to aid radiologists in identifying and classifying lesions while also reducing reporting time by increasing the quantitative assessment of the disease. In fact, prostate CAD with MRI has drastically evolved during the past 10 years. Machine learning (ML), a branch of AI, and its sub-discipline, deep learning (DL), have become popular in medical imaging because of their ability to understand vast amounts of data. These technologies are providing radiologists with many features based on these magnetic resonance prostate images like segmenting the prostate, a registration of the different images, extracting features, detecting lesions and classifying the lesions as cancerous or non-cancerous. Recent studies were also able to predict the Gleason Grade Score based on MRI and radiomic analysis [6] [7].

The goal of this Master Thesis is to build a full deep learning based CAD pipeline that would be used by doctors and radiologists to not only improve their diagnosis but to also assist them in their screening of prostate cancerous lesions by providing a clear view on what the learning sees and detects. The built CAD should take as input a patient's mp-MRI results and would output a segmentation of the prostate as well as the detected and classified lesions, visualised using heat maps that would help the radiologists in their assessment of the images.

1.2 Related Work

Many investigations have been done to try and build a machine learning model that could help doctors with prostate cancer diagnosis. These CAD systems usually target different goals, from prostate segmentation and volume approximation, to lesion aggressiveness prediction. In this section, we will present an overview of the various models developed to deal with three aspects of computer aided diagnosis for prostate cancer : prostate segmentation, lesion detection and lesion classification ([8] was used as a basis of inspiration to write this section, as well as some of the below mentioned papers).

1.2.1 Prostate Segmentation

MRI images usually capture a large zone around the organ that is aimed at. In terms of prostate mp-MRI, the different images that are obtained show the prostate gland with its surroundings. Radiologists usually segment the prostate to be able to focus on the important parts of the different images. Besides being used for research purposes, a segmented prostate in an image could greatly improve the study of the prostate's growth (volume/dimensions computation), which is clearly an indication to different prostate diseases as seen previously.

Prostate segmentation is usually performed in a manual or semi-automated fashion and is limited by inter-observer variability [8]. To meet this need of better and faster prostate segmentation, various ML models have been developed. Fuzzy C-Means (FCM) is a clustering algorithm where each data point is assigned a probability score to belong to a certain cluster. It is an unsupervised machine learning technique that was used by Rundo et al. [9] to process T1-weighted and T2-weighted prostate MRIs and segment the prostate gland (a description of the different MRI modalities will be presented in Chapter 2). Their study evaluated 21 patients to achieve an average Dice Similarity Coefficient of $90.77 \pm 1.75\%$. The Dice Similarity Coefficient (DSC) also known as the Dice Coefficient is a statistical tool which measures the similarity between two sets of data. In image segmentation, it is commonly used to assess the spatial intersection between two images (scored from 0 to 1 or 0 to 100). Rundo et al. also concluded that combining these two types of MRI significantly enhanced the prostate gland segmentation results, compared to using one of the two types.

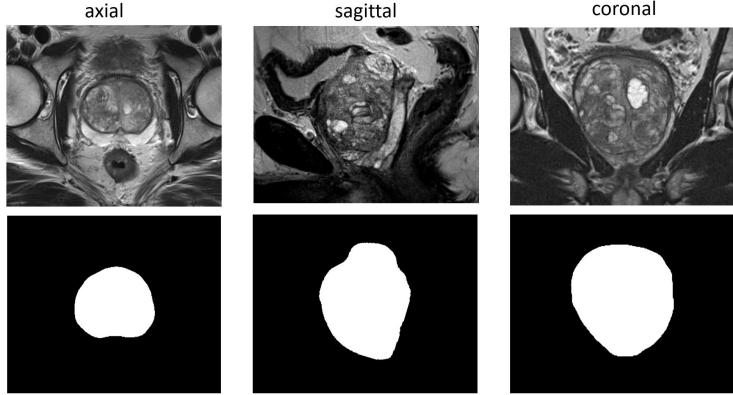


Figure 1.2: Prostate Segmentation via the 3 planes [3]

Even though machine learning models achieved good scores, deep learning approaches were also researched and used for this task, since they accomplish feature extraction independently. Zhu et al. [10] built a cascaded fully convolutional network to accomplish the prostate segmentation task. Their purpose was to develop a fully automatic approach to not only to segment the prostate from the rest of the image, but to also segment the outer contour and the peripheral zone of the prostate [10]. They used two cascaded U-Net networks that they trained on two types of MRI images : T2-weighted and Diffusion-weighted MRI (a description of the different MRI modalities will be presented in Chapter 2). U-Net is a U-shaped fully convolutional network architecture proposed by Ronneberger and Olaf et al. in 2015 [11]. It successively compresses an image, derives features during these contractions, and classifies every pixel in the image. On a population of 163 subjects, including 61 healthy subjects and 102 prostate cancer patients, they obtained a Dice Similarity Coefficient (DSC) of $92.7 \pm 4.2\%$ for the total while prostate gland and $79.3 \pm 10.4\%$ for the total peripheral zone on testing sets [11].

Milletari et al. used a similar approach to the U-Net network, but upgraded the algorithm to enable it grasping 3D volumes [12]. In fact, most convolutional Neural Networks are only able to process 2D images, while MRI images are of three dimensions. They named their model V-Net. Their model was trained end-to-end on prostate MRI volumes and segments the whole volume at once, achieving a DSC of $86.9 \pm 3.3\%$ on testing sets after being trained on a 50 MRI volumes dataset. Another study done by Yan et al. [13] investigated a Propagation Deep Neural Network model. This model that incorporates the optimal combination of multi-level feature extraction as a single model, is based on 3 types of connected layers : convolution and pooling layers, a propagation layer and a loss measure layer. By training their model on 50 patients and testing it on 30 patients, they achieved a competitive $84.13 \pm 5.18\%$ DSC on testing sets [13].

To sum up, we have seen that different machine learning models are being developed to segment the prostate in MRI slices (2D) and volumes (3D). We can conclude that these models could lead to very accurate results that could definitely be used for PCa screening and diagnosis.

1.2.2 Lesion Detection

Once the prostate is localised by a computer or a doctor, the next step would be to detect the regions in the gland that could potentially be dangerous. We call these regions 'Lesions'. A prostatic lesion is an abnormal growth of cells that may or may not be significantly cancerous. Detecting these regions of interest will not only be useful for our next step which is judging if these lesions are cancerous or not, but it could also potentially help the doctor visualising the different abnormal zones in the prostate and so help with the diagnosis.

As we have said, doctors are often prone to error in detecting these lesions due to different reasons (lack of experience, fatigue, a lot of variability ...). Using a machine learning approach could potentially be beneficial here since it would study the different patterns in the images between prostatic lesions. Lay et al. [14] proposed a machine learning CAD approach based on random forest to detect these lesions. The method uses a combination of spatial, intensity, and texture features extracted from mp-MRI sequences (T2W, ADC, and DWI). Note that a description of the different MRI modalities will be presented in Chapter 2). Their study was done on a dataset of 224 patients for a total of 287 benign lesions and 123 cancerous lesions. Their ML technique lead to an area under the curve (AUC) score of 0.93. To define AUC, it is a measurement for binary classification and ranges from 0 to 1. This amazingly high result obtained shows that a machine learning model can definitely help detecting prostate lesions with high accuracy.

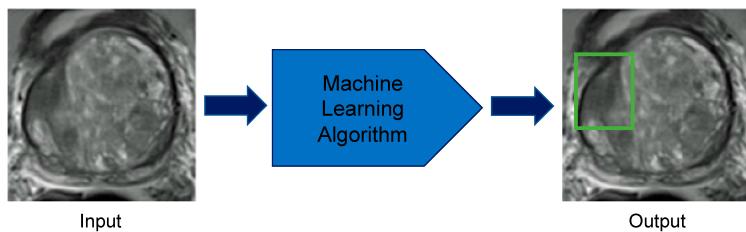


Figure 1.3: Input and Output of a Lesion Detection algorithm [8]

As for prostate segmentation, different deep learning techniques were also developed for lesion detection. Sumathipala et al. [15] built deep convolutional neural networks with a goal to generate probability maps overlaid on original mp-MRI images to help radiologists identify suspected abnormal zones. They constructed this CAD by optimising the holistically nested edge detection (HED) deep convolutional neural network, which is a neural network architecture that was first created for detecting edges in natural images and has since increased in popularity for use in CT scan anatomy segmentation [16]. Their study dataset consisted of 186 patients (T2W, ADC and DWI images were acquired for each patient) from six different institutes. They achieved an AUC of 0.97 ± 0.01 .

Tsehay et al. [17] also used a deep learning algorithm based on a convolutional neural network architecture. They built this architecture from a state-of-the-art edge detector that takes an image as input and produces an image probability map as output. Their study consisted of 52 patients where mp-MRI was acquired for each patient (T2W, ADC

and DWI). They were able to achieve an AUC of 0.90 which is another proof that these models have a non-negligible potential in this field as they are reaching very high accuracy.

Xu et al. tried solving this problem by using Residual Networks [18]. 346 patients were used for their study (T2W, ADC and DWI). They reached an impressive AUC of 0.97.

The high accuracy values obtained via the studies above, show that prostate lesion detection is a task that could be very well managed by AI models. Of course, a lot of research has still yet to be done to obtain a usable and accurate enough CAD, but the combination of machine and human intelligence could potentially lead to the best results in term of detection and diagnosis.

1.2.3 Lesion Classification

In the two previous parts, we saw how different algorithms could segment the prostate gland in prostate mp-MRI and how they could detect different lesions in these images. As we know, lesions aren't always cancerous, some lesions are significant (cancerous) while some are not at all (not significant) resulting in unnecessary biopsies or treatment. Now let us see what research has been done in this field, to not only detect lesions but to try and predict if these lesions are cancerous or not.

Liu et al. [19] developed a novel deep learning architecture, that is known as *XmasNet*, to try and solve this problem. Their convolutional neural network (CNN) based architecture was trained and tested on the mp-MRI data provided by the **PROSTATEx** [20] challenge (341 cases with 241 for training with T2W, ADC and DWI images for each). They achieved an AUC of 0.84 which was one of the highest scores of this particular challenge. By building their novel model, they showed that deep learning has a very high potential in completing the task of classifying prostate lesions.

Kiraly et al. [21] also built a deep learning model, but used a completely different model based on encoder-decoders. They proposed a multi-channel, image-to-image, convolutional encoder-decoders, where responses signify localized lesions and output channels represent different tumor classes. Compared to the previous algorithm, by encoding and decoding, this technique allows for both localization and classification within a single run [21]. This approach is very similar to the U-Net network which was presented earlier on. They also used the **PROSTATEx** dataset where they were able to achieve an AUC score of 0.83. Even though if this score is similar to the result of the *XmasNet* model, this study shows us that using an encoder-decoder model could be truly beneficial and could generate good results for this task (the advantage in using an encoder-decoder model is the fact that segmentation and classification can be done in a single run).

Liu et al. [22] developed a two-stage approach to not only classify lesions but to also detect them, within prostate mp-MRI. The two stages consist of a Mask R-CNN stage and a Weakly Supervised Deep Neural Network stage. R-CNN [23] is an image segmentation model that smartly goes over the image with different sized rectangles and runs

each region through a convolutional neural network to generate a feature vector representing this region. Mask R-CNN [24] is an extension of Faster R-CNN [25] that differs by adding a branch for predicting an object mask in parallel. Liu et al. trained in the first stage of their approach a Mask R-CNN model to automatically segment prostate structures. In the second stage, they built a weakly supervised deep neural network to detect and classify lesions in a single run [22]. They used two different datasets to train, test and validate their approach : the previously mentioned **PROSTATEx** challenge dataset and another dataset from their local cohort (images used : T2W, ADC and DWI). They obtained an AUC score of 0.912 on the **PROSTATEx** dataset and 0.882 on the other one.

Transfer learning is a machine learning technique in which a model created for one job is utilized as the basis for a model created for a different task. This method is typically used in situations where the learning of a new task relies on a previously learned task. This generally leads to a faster process and requires less training data for the second task. Yuan et al. [26] built a novel multi-parametric magnetic resonance transfer learning (MPTL) method to automatically classify prostate cancer. Their model consists of two parts, the first is a three branch architecture which transfers a pre-trained model to compute features from different mp-MRIs (T2w transaxial, T2w sagittal and ADC). The second a network that concatenates the obtained features to classify the different regions. They trained their model on two datasets : an institutional dataset containing 132 cases, and they also took 112 cases from the **PROSTATEx-2** challenge [7]. The model achieved an AUC score of 0.896.

To conclude this overview of the different studies on lesion classification, we can say that different methods are successfully being developed to classify detected lesions in the prostate. Many of these methods rely on machine learning and specially deep learning. We can also see that lesion detection and classification are two tasks that can often be merged in one. In fact, it could be useful to detect the lesions without directly classifying them, which would give more freedom to the radiologist's input, but it can also be very useful to classify the lesions, which would give the radiologist a "machine opinion" that he could base himself on before completing his diagnosis. We can also conclude from the precision of the above mentioned methods, that they are reliable in terms of CAD and could potentially generate even better diagnosis than current radiologists.

1.2.4 Grad-CAM

Grad-CAM is a technique introduced by Selvaraju et al. [27] for producing "visual explanations" for decisions taken by convolutional neural networks. In fact, when using a neural network for detection, segmentation, classification or any other image processing task, it is quite hard to dive into the connections inside the network and understand how the network grasps the images and analyses it. Gradient-weighted Class Activation Mapping (Grad-CAM), generates a coarse localization map highlighting important regions in the image depending on the predicted concept, and does that by using the gradients flowing into the final convolutional layer of a network that predicts this target concept [27]. These features generated by Grad-CAM help us understand what the network "sees" and how the neurons are analysing the image for the particular task achieved by the model.

This technique isn't specially used for prostate segmentation or lesion detection/classification, but it is a really useful method to adopt when dealing with any kind of image segmentation task via deep learning. It is being introduced here because it is part of the related work done in image processing and artificial intelligence and will be used later on in this research to provide better visual insights to the user.

1.3 Research Objectives

The world is turning digital. Our phones, our cars, our money, our schools ... Almost everything we know is currently governed by digital devices, and that is because technology and engineering are making life easier through these digital instruments. As humans, it is in our nature to always seek for the easiest and most efficient solutions to our problems. Artificial intelligence is not only helping us finding these solutions, but it is doing it in a faster and better way than we are. From recommending you with shopping items that you didn't know you needed before a machine learning algorithm showed them to you, to unlocking your phone using your face or voice, AI is not here to take part with human intelligence; it is here to take over.

Even though our hospitals don't have yet self walking robot nurses, we can clearly see from the different studies presented above that AI and DL have a very high potential in producing accurate diagnosis, even better than current doctors. These algorithms are not only able to take in a load of information that is much higher than usual doctors, but they can also deal with different images at the same time as well as see detailed features that the human eye isn't capable of grasping.

AI in the medical field is evolving, and even though we are not at a point where it could be replacing a diagnosis from a real doctor, it has a high potential of taking over within the next few years. But most humans, and specially engineers and doctors fear these models. First and foremost, even if these algorithms could produce very accurate results, engineers have a clear idea on their flaws and the fact that machine "intelligence" is still very far from human intelligence. Second, doctors fear for their jobs, they fear that AI will take over and the world won't necessarily need as many doctors as we currently need. But the question that most of us scientists ask ourselves today is : if you were to be a patient, would you prefer getting diagnosed by an algorithm or a doctor? Maybe both?

In front of AI's revolutionary evolution, two choices are possible : either to stand still and let it take over or to form a human-machine synergy that could result in the best of both worlds. The goal of this research is to build a computer aided diagnosis software designed for a typical doctor-machine synergy. The software would potentially be used by radiologists or other doctors to better the diagnosis of prostate cancer via a non-invasive technique, and that is by analysing mp-MRI images. The goal of the software is not to produce a final diagnosis of a patient's prostate cancer state (which would be AI taking over the diagnosis), but to produce insights and visualisations that would help the doctors in generating easier, faster and more accurate diagnosis. These insights are in the form of : prostate segmentation, lesion detection and lesion classification. To recapitulate, here are the 3 main objectives of this research :

1. Work on three axes : prostate segmentation, lesion detection and lesion classification. The goal of this master thesis is to build a software that will not only work on a specific theme in prostate cancer detection, but all three. The goal isn't to perfect a single task, but to try and merge all three in a single research

2. Provide insights to help doctors in generating faster and more accurate diagnosis. The built software won't generate a complete diagnosis that would replace the doctor's diagnosis, instead it is designed to be used by doctors in order to generate a faster and more accurate prostate cancer diagnosis
3. Enhance Human-Artificial Intelligence synergy. This research aims to work towards a better synergy between human intelligence and machine intelligence by providing a software that will be used by doctors and will assist them by providing visual insights which would help them in their diagnosis of prostate cancer

This master thesis will be divided in three main chapters. The different datasets as well as the materials and methods used in each stage of the CAD pipeline will be presented in the first part of the following chapter. A second part of this chapter will be dedicated to the deep learning models, architectures and technicalities that were built and used. In chapter 3, the different experiments and results that allowed the validation and testing of the software will be proposed, before ending this thesis with a conclusion and further improvements chapter.

Chapter 2

Materials and Methods

To be able to build, train and test any machine learning model, a valid dataset and a studied architecture are needed. In this chapter, the datasets used to build the different models will be presented as well as the architecture of each and everyone of these models. All the details corresponding to the implementation of the models will be given, as well as the pre-processing steps so that the entire pipeline could be reproduced by any reader. This chapter is divided in three parts :

1. Firstly, an overview of the different MRI imaging techniques that were used in the different prediction models. The goal of this overview is to familiarise the reader to the different types of MRI images used for prostate screening
2. Secondly, a description of the different datasets that were acquired to train and test the models for both parts of the pipeline : segmentation of the prostate and the detection/classification of the lesions within the prostate gland
3. Lastly, a detailed presentation of the models' architectures, as well as the different steps that were followed from pre-processing the data to training and validating the models

2.1 MRI Imaging Techniques

As the goal of this research is to build a computer software based on Magnetic Resonance Imaging, to help doctors with the diagnosis of PCa via this non-invasive method, it is crucial to present an overview of the different MRI techniques acquired for prostate cancer detection and diagnosis. In fact, each modality has its own features that indicate the presence of abnormal cells in a particular way. This section serves as an overview of the different imaging techniques that will be used for the training of the models, as well as a way to show the reader that mp-MRIs can be used to detect cancerous prostate regions but with some difficulties and ambiguities. Note that other techniques that the ones presented below are available, but won't be presented due to their lack of usage for the particular task of PCa detection.

The magnetic characteristics of atomic nuclei are used in MRI. The randomly oriented protons inside the water nuclei of the tissue being investigated are aligned using an

external magnetic field. An external source of Radio Frequency (RF) radiation is subsequently used to disturb this magnetization. The nuclei return to their resting alignment via various relaxation mechanisms, generating RF energy in the process. After a specified period of time has passed from the first RF, the emitted signals are calculated. The Fourier transformation is used to convert the frequency information in the signal from each place in the scanned plane to intensity levels, which are then represented as shades of gray in a matrix of pixels. Changing the sequence of RF pulses applied and captured can produce a variety of images [28] [29].

Repetition Time (TR) is the amount of time between successive pulse sequences applied to the same slice. Time to Echo (TE) is the time between the delivery of the RF pulse and the receipt of the echo signal [30].

T1 and T2 relaxation times are used to describe the tissues of the human body that are investigated using MRI [29]:

- The pace at which stimulated protons return to equilibrium is measured by T1, or longitudinal relaxation time. It's the amount of time it takes for spinning protons to re-align with the external magnetic field
- The pace at which stimulated protons attain equilibrium or move out of phase with each other is measured by T2, or transverse relaxation time. It is the time it takes for spinning protons to lose phase coherence with nuclei spinning perpendicular to the main field

2.1.1 T2-W-MRI

T2 Weighted MRI is one of the first MRI-modality used to diagnose anomalies in the body using MRI. Nowadays it is fairly used for the detection and localisation of anomalies within the prostate gland. This imaging technique is very well suited to render the zonal anatomy of the prostate, as it gives the clearest overview of the gland and its different zones. T2 Weighted images are produced using long TE and TR times, which leads to the domination of the contrast and the brightness of the images by the T2 properties of the tissue. In terms of PCa, it is characterized by a low signal intensity (SI) on the T2-W-MRI in the peripheral zone and having a round or ill-defined shape. In the central gland, PCa is also characterized by a low SI but its shape is more homogeneous which makes it more difficult to detect since healthy tissue has also a low SI in the central gland, compared to a high SI in the peripheral zone. These ill-shaped masses detect on T2-W-MRI are also characteristics of BPH and other prostate diseases, which makes it even harder for radiologists to assess if a particular zone translates to a cancerous one or not [4] [28].

2.1.2 DCE-MRI

Dynamic Contrast Enhanced MRI is a perfusion MRI technique that exploits the vascularity characteristics of tissues. The technique consists in injecting a contrast bolus (gadolinium based) intravenously in the patient, and obtaining rapid T1 images. The

signal received is related to the concentration of gadolinium which in turn depends on a number of factors (intravascular gadolinium, permeability, surface area, blood flow ...) which are of course affected by abnormal cells. K-trans is the most commonly computed parameter via this MRI technique. It is a measure of capillary permeability and is computed by measuring the accumulation of gadolinium-based contrast agent in the extravascular-extracellular space. Abnormal zones are typically characterised with a higher SI in K-trans images [4] [28].

2.1.3 DW-MRI

Diffusion Weighted MRI is a modality that exploits the variations in the motion of water molecules in different tissues. The diffusion movement of water molecules is typically free within the prostate healthy tissues. PCa growth leads to the destruction of normal glandular structure and its associated with an increase in cellular density. These factors lead to a decrease in the diffusion of water molecules as they cause a restriction on these movements. It is on the basis of this principle that PCa could be detected in DW-MRI [29].

The b-value is a factor that reflects the strength and timing of the gradients used to generate DW-MRIs. Having measured a set of at least 2 different b-value images (e.g., b 0 and b 1000 s/mm²) the system calculates pixel by pixel the apparent diffusion coefficient (ADC) by linear regression. This coefficient is a measure of the magnitude of diffusion of water molecules within the tissue. The ADC values per pixel are then combined together to form the ADC map. This map could come in use since prostate cancer appears as a low SI region in this map, compared to healthy tissue which usually appears as high SI [4] [28].

2.1.4 PDW

Proton Density weighted MRI reflect the density of protons in a tissue (number of protons per volume). It is an intermediate sequence that shares some features of both T1 and T2. Tissues with few protons have a low SI compared to tissues with a high density in protons. It is an imaging technique that is mostly used in brain MRI to evaluate gray/white matter pathology [31].

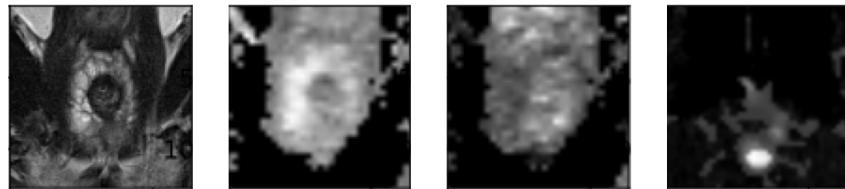


Figure 2.1: Prostate MRI modalities. From left to right : T2W, ADC map, B-Val image, K-trans [20]

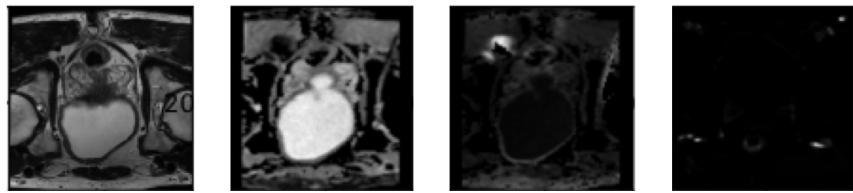


Figure 2.2: Prostate MRI modalities. From left to right : T2W, ADC map, B-Val image, K-trans

As we can see from the provided description and images, mp-MRIs have an enormous potential in helping doctors detect cancerous zones in a patient's prostate. The different problems and uncertainties arise from the fact that multiple images and types of images need to be visualised and analysed by the doctors in order to output a valid diagnosis. This large quantity of data can be overwhelming for the human eye and brain, as well as the details that it contains. Another constraint is the visibility and the experience of the observer. As we have seen, the appearance of cancerous zones really differ depending on their location within the prostate, plus the variation in the contrast can sometimes be invisible to the human eye, specially in the case of an inexperienced doctor. For all these reasons and the ones mentioned above, it is clear that deep learning could have a non-negligible potential in helping doctors detect prostate cancer via mp-MRIs.

2.2 MRI Datasets

After a brief introduction to different MRI modalities, this section contains all the information about the different MRI datasets that were used to train, validate and test the deep learning models.

2.2.1 Prostate Segmentation

Prostate segmentation in MRI scans has always been a challenge to automate via learning models. The **PROMISE12** challenge [32] was organised by the Medical Image Computing and Computer Assisted Intervention Society (MICCAI) in 2012 with a goal to compare interactive and (semi)-automatic segmentation algorithms for MRI of the prostate. Being a well-recognized benchmark for prostate MR image segmentation within the scientific community, the dataset consists of 50 training cases with corresponding manual masks and 30 testing cases from different centers and MR scanner vendors (see table 2.1). Each case has 15 to 54 transverse T2W prostate MR images. Note that the labels of the testing cases are not supplied, since they are meant to evaluate the challenge algorithms. The repartition between training and testing data will be discussed later. In the table below [33], we can see the details of the acquisition protocols for the **PROMISE12** dataset [32]. The cases were acquired from four different centers around the world [33]. Each center provided a total of 25 cases, which 20 of them were only presented during the live version of the challenge and not available for download from the challenge's website [32].

Center	Field Strength (T)	Endorectal Coil	Manufacturer
HK	1.5	Yes	Siemens
BIDMC	3	Yes	GE
UCL	1.5 and 3	No	Siemens
RUNMC	3	No	Siemens

Table 2.1: Information about the different MRIs provided from the four institutes in the dataset [33]

The dataset is completely available for download [32] on the challenge website. This dataset is divided in training and testing images. For the training images, corresponding prostate masks are also given. For the testing images these masks are not give, since it is a challenge, the goal was to predict the masks on the given test set and compare the results of the different teams that participated.

After downloading and doing some analysis on the data, here are some results that represent the usable images in this dataset :

- Number of patients : 50 usable cases
- Each patient has an MRI volume, these volumes have different shapes depending on the patient (different imaging techniques, acquisition ...) :

- Each volume contains a number of slices varying between 15 and 54, with a mean of 22.7
- Each slice has a shape of 320x320 or 512x512
- 1377 total usable images (to be divided in training, testing and validation)
- From these 1377 images, 778 contain a prostate and 599 don't

2.2.2 Lesion Detection and Classification

For the task of detecting the lesions in the prostate and classifying them as cancerous or non-cancerous, an online challenge was also used as a dataset provider. In fact, the availability of accurate and clean datasets providing prostate mp-MRI scans is considerably low. The dataset of the **PROSTATEx** challenge was used in this part [20]. This competition was held in conjunction with the 2017 SPIE Medical Imaging Symposium and its main goal was to predict the clinical significance of prostate lesions found in MRI data [20].

The types of MRI available in this dataset are the following : T2-weighted (T2W), proton density-weighted (PDW), dynamic contrast-enhanced (DCE), and diffusion-weighted (DWI) images. The images were acquired on two different types of Siemens 3T MR scanners : the MAGNETOM Trio and Skyra [20]. T2W images were acquired using a turbo spin echo sequence and had a resolution of around 0.5 mm in-plane and a slice thickness of 3.6 mm. The DCE time series were acquired using a 3-D turbo flash gradient echo sequence with a resolution of around 1.5 mm in-plane, a slice thickness of 4 mm and a temporal resolution of 3.5 s. The DWI series were acquired with a single-shot echo planar imaging sequence with a resolution of 2 mm in-plane and 3.6 mm slice thickness and with diffusion-encoding gradients in three directions. It is important to note that all images were acquired without an endorectal coil [20].

The Apparent diffusion coefficients (ADC) were also computed from the DWI scans and provided along the B-values (50, 400 and 800) that were used to compute these coefficients. K-trans images were also provided [20].

To recap, each patient has one study with several MRI images (T2W sagittal transverse and coronal, ADC and Bval computed from DWI, PDW and DCE) and one Ktrans image. The dataset counts 344 participants which of whom 204 were dedicated for training (available significance result) and 140 for testing and evaluating the challenge score. The data can be downloaded from the Cancer Imaging Archive website [20] and are provided in the *DICOM* format except the Ktrans images which come in the *mhd* and *zraw* format [20].

After downloading and doing some analysis on the data, here are some results that represent the usable images in this dataset (by usable it is meant here the data that can be used for training and testing. In fact, the dataset contains much more images than what will be described below, but not all these images are usable, which induced

a necessary reduction in the number of usable images. More details on how the number of images dropped in chapter 2.3.3). It is important to mention that here, we count the number of images per type. So 1 image here represents multiple images, one for each type of MRI provided :

- Number of patients : 204 usable patients
- From these patients, we have 224 cases of non-significant lesions and 77 cases of significant lesions (a patient could have multiple lesions)
- For each volume of MRI (each patient has a volume for each type) :
 - We have a number of slices ranging from 18 to 27, with a mean of 20.42
 - Each slices has a shape of 320x320, 384x384 or 640x640
- A total of 4201 images are available
- On this total number of images :
 - 855 images contain nothing but surroundings (no prostate or lesions)
 - 2067 only contain a prostate (no lesions)
 - 1279 contain lesions
 - From this 1279, 858 contain non-significant lesions and 332 contain significant lesions. 89 contain both

From these metrics we can already see that the number of images is relatively low. From all the usable images around 20% are black images, and around 50% only prostate. These images could be used for training but create imbalance in the data, since the goal is to detect, and classify lesions. It is also important to take into consideration the difference between the number of images that contain non-significant lesions and significant lesions. All these differences will help understand and evaluate the results generated by the foreseen built model.

2.3 Deep Learning Pipeline

As the goal of this research is to build a software that would essentially accomplish prostate segmentation and lesion detection/classification, it is necessary to provide the reader with the different steps that were followed from acquiring the previously mentioned datasets to training and validating both models. This section will start by an introduction to the U-Net architecture which is the basis of both models, to follow up with the previously mentioned steps for each task and to finish off with a last section part, dedicated to an overview of the entire software pipeline : from the radiologist's input, to the machine's output.

2.3.1 U-Net Architecture

U-Net is a convolutional neural network that was precisely developed for the task of biomedical image segmentation [11]. A convolutional neural network (CNN) is a category of deep neural networks that is mainly used for image analysis (it contains convolutional layers which are based on the convolution operation, thus the name). CNNs are very useful for classification tasks, where the output to an image is a predicted class label. However, typical CNNs are not made for a pixel-based classification, or assigning a class to every pixel in an image (which is basically segmentation). This could be very useful since classifying pixels translates to a certain localisation in the output. Image segmentation could be achieved using typical CNNs but it would require some sort of sliding window, not only does this slow down the process, but it also introduces a trade-off between localization accuracy and context utilization.

Researchers at the University of Freiburg in Germany, Ronneberger et al. aimed to solve that problem by building a network with an architecture that consists of a contracting path to capture context, and a symmetric expanding path that enables precise localization [11]. On the image 2.3 a visualisation of the basic U-Net architecture is demonstrated. They showed in their work that this particular form of network could be trained end-to-end on image segmentation, from very few images, and outperforms other deep learning based methods [11].

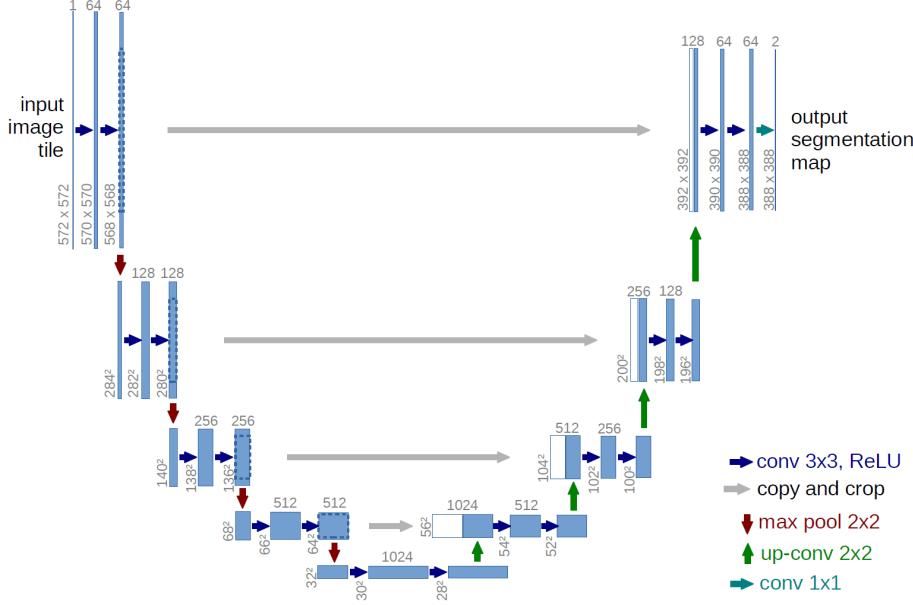


Figure 2.3: U-Net architecture [11]

As seen on the image, the architecture consists of a contracting path and an expansive path (whereas the "U" letter in U-Net). To fully understand the mystery behind these two paths, here are some definition that will be useful in understanding each step of the architecture [34] [35] :

- Convolution operation : takes an image as input, convolves it to a set of filters, then outputs a 3D volume called feature map. The convolution operation computes the sum of the element-wise multiplication between the input image and the kernel
- Max Pooling operation : A pooling function's purpose is to reduce the size of a feature map so that we can work with fewer parameters. The aim is to keep only the most interesting aspects of each zone (max valued pixels for example) and throw away the information that is judged as "not as important". Note that this operation reduces the size of the image, as it can be called *down sampling*
- Up-sampling : as the previous operation is meant to down sample an image, there exist operations to up sample images. These functions convert the obtained low resolution image to a higher resolution in order to recover the localisation information in the image
- Transposed convolution operation : this operation is a typical example of an up-sampling method. It's used to up-sample an image using learnable parameters [36]. Its the backward operation of the basic convolution operation, where the weights of the convolution matrix are learnable

Now that the building blocks of this architecture have been presented, here is a detailed explanation on both paths of the U-Net model [34] [37] [35] :

- Contracting path : in this branch of the architecture, the goal is to extract features from the image. This is done by applying a series of convolution operations and max pooling. Note that here we loose any kind of spacial information, but we gain in feature maps (loose the *where*, gain the *what*). The contraction takes as input an image (in the paper it is a 572x572x1 shaped image). Then it applies two convolution operations to obtain a 568x568x64 shaped feature maps. A max pooling operation follows to down sample the size of the feature maps to 284x284x64. The contraction path then iterates over these two operations to finally obtain 28x28x1024 shaped feature maps
- Expansive path : here the main operation used is the transposed convolution for up-sampling. The parameters for each transpose convolution are such that the height and width of the image are doubled while the depth is reduced in half. The goal of this path is to recover the spacial information (regain the *where* based on the computed *what*). To help with the recovery of the spacial information, at every step of the decoder, skip connections are used by concatenating the output of the transposed convolution layers with the feature maps from the encoder that are at the same level. After each concatenation, we add two convolutions in a row so that the model can learn to put together a more precise performance. So the followed operations at every step of the way are : up-sampling, concatenation, convolution two times

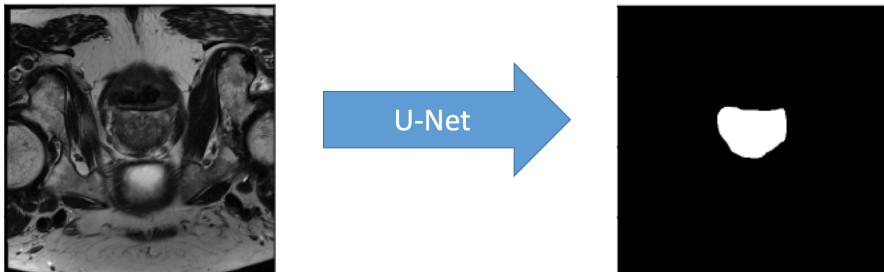


Figure 2.4: Segmentation of a prostate gland on a T2-W MRI axial view using a U-Net based model [20]

2.3.2 Prostate Segmentation

In this section, the different steps that were taken to build, train and validate the model to segment the prostate will be presented.

Pre-processing

As mentioned previously, the data made available by the **PROMISE12** challenge already contained the T2-W-MRIs and the corresponding prostate masks, which facilitated the work since not much pre-processing was required. Since any deep learning network

takes as input same sized samples, a first step to pre-process the data is to set all the images and masks to the same shape, which was chosen here to 256x256. In fact the images made available by the challenge had different sizes depending on the patient (varying between 320x320 and 512x512) so a standard size had to be chosen for all images. When a standard image size has to be chosen, what is usually taken into consideration is the processor and the memory of the computer that is meant to hold the training and testing of the model on the chosen dataset. In the case of this research, different processors were used to run different tests, so the choice of the standard image size couldn't be directly correlated to a specific computer. The standard size of 256x256 was chosen due to its popularity on the internet [38] and the fact that the online repository that was used as a basis inspiration for this part of the research used a standard size of 256x256 [39](this repository will be introduced in details later on). A standard size of 256x256 makes also sense in terms of the number of parameters in the deep learning model, since it is always smart to start with a low amount of parameters and then augment if needed.

Once that is done, the images are all stacked over one another in one *Numpy* array, as well as the masks. This results in two concatenated *Numpy* arrays each of dimension Nx256x256 where N represents the number of images. After that, a curvature driven image denoising algorithm is applied to smooth the different T2-W images. The advantage of this approach is that sharp boundaries are preserved with smoothing occurring only within a region [40], which could potentially help out with the segmentation task. To finalise the pre-processing step, normalisation is applied to the the MRI images by computing the mean and the standard deviation of the training set, then for each set (training, validation and testing) subtracting from it the mean and dividing it by the standard deviation of the training set.

Network Structure

The scope of the network is to take as input a T2-W-MRI (axial view) and to generate a mask (black and white image) where the white pixels indicate the prostate gland. So the network takes in a two-dimensional image and outputs a two-dimensional mask. The architecture of the model is based on the U-Net architecture presented previously, with some particularities mentioned below :

1. The number of channels in the first convolutional layer is 8 (compared to 64 in the original paper)
2. The network has a depth of five. The depth of the network can be seen as the number of operational blocks within each branch. More specifically, it could be represented by the number of max pooling operations done in the contracting path. In the original paper (see figure 2.3), the image input shape is 572x572, and with a depth of 4 the image obtained at the bottom of the network has a shape of 28x28. We can see that the shape was reduced by a factor of 20. If we apply the same logic in the network that built for prostate segmentation, the image input shape is 256x256 and $256/20 = 12.8$. Since the numbers of channels for the first convolution is 8, the image shape will reach a factor of 8. Between a shape of 16x16 for a depth of 4 or a shape of 8x8 for a depth of 5, a depth of 5 was chosen. A depth of 4 could

also yield good results and be correct, but since the goal of this master thesis isn't to test every possible parameter and obtain the best model, it was chosen to go for a depth of 5 and in case of none satisfying results, test the depth of 4

3. Zero padding and Rectified Linear Unit are used for the convolutions
4. *Sigmoid* is used as an activation function for the last layer
5. Variance scaling is used for kernel initialisation. To give a bit of context on the notion of kernel initialisation, *Xavier Initialisation* is a method that was proposed by Xavier et al. [41] which is a weights initialisation technique that tries to make the variance of the outputs of a layer equal to the variance of its inputs. This method usually yields faster convergence in terms of learning. Variance initialisation is an initialisation method similar to the *Xavier Initialisation* that is generalisable for ReLu (compared to the previously used sigmoid activation)
6. After each convolutional layer, batch normalization is applied to make the learning faster and more stable
7. Between two convolutional layers, a dropout of 50% is used to regularize the network and avoid overfitting.
8. Residual connections are also used within the convolutional blocks. Residual connections are very similar to 'skip connections' [42]. Non-linear activation functions can cause the gradients to explode or vanish, which is where skip connections can have an effect. They form conceptually a 'bus' which flows right the way through the network and in reverse, allowing each block of network layers to tap the values at a point along the bus and then add/subtract values onto the bus. This allows the gradients to flow through a network directly, without passing through non-linear activation functions [42].

To recap, the network built to segment the prostate within T2-W-MRI is based on the U-Net model presented previously, with some added particularities : the input image size is 256x256, the depth of the network is 5 (5 max pooling operations), padding is added to help avoid the loss of information on the corners, variance scaling was added for kernel initialization and finally residual connections are used between convolutional blocks. Finally, the model takes as input a T2-W-MRI of shape 256x256 and predicts a mask of shape 256x256 where the white pixels indicate the segmented region.

Training & Validation

As previously mentioned, the dataset consists of 50 patients. Each case has a T2-W-MRI volume associated to it. Since every volume (3D) is composed of many slices (2D) and since our segmentation is done in a two-dimensional plane, we obtain in total 1377 images with corresponding masks. Since the challenge didn't provide the corresponding masks to the test set, we need to train validate and test the model on the dataset's training cases (50 cases in total). We split these cases so that 15% goes to testing, and 85% to training which leads to 8 cases for testing and 42 for training (we separate the cases and not the images so that no correlation between the testing and training data could be

present). The training set is then also split into training and validation. For validation, we randomly choose before compiling the model, 5 cases between the 42 that would serve as validation cases.

For training the model, the loss function used is based on the Dice Similarity Coefficient. This coefficient was already introduced in the first chapter when talking about the previous work done in prostate segmentation as it is suitable for image segmentation evaluation. The DSC ranges from 0 to 1 and is computed via the following formula in both images that ought to be compared :

$$(2 * \text{Area of Overlap}) / (\text{total number of pixels in both images}) \quad (2.1)$$

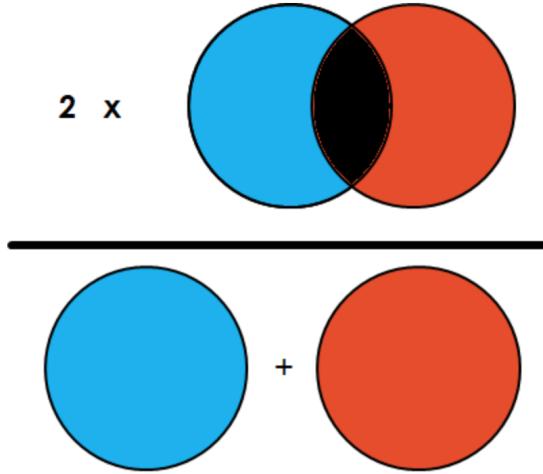


Figure 2.5: Informative Figure of the DSC formula [43]

As a result, the DSC is a measure of mask overlap. A complete overlap is indicated by a 1; no overlap is indicated by a 0. The DSC can be used as a metric for evaluating the performance of the model and the Dice Loss which is equal to $1 - \text{DSC}$ can also be used as a loss function during training. To be able to use the Dice Loss as a loss function, some modification needs to be done. In fact, the DSC is not differentiable since it is based on a binary input and intersection. So a gradient can't be computed out of it, which prevents it from taking the form of a loss function. To be able to do that, a Soft Dice loss needs to be implemented, that directly uses the predicted probabilities instead of thresholding and converting them into a binary mask. The main difference in the formula between the Soft Dice Loss (which is usually called the Loss Dice) and the formula presented above which is the formula for the typical DSC can be seen below, where for the Soft Dice Loss we have a multiplication on the numerator instead of an intersection :

$$\text{Soft Dice Loss}(y, \hat{y}) = 1 - \frac{2 \sum_i^N y_i \cdot \hat{y}_i}{\sum_i^N y_i + \sum_i^N \hat{y}_i} \quad (2.2)$$

$$DSC = (2 * \text{Area of Overlap}) / (\text{total number of pixels in both images}) \quad (2.3)$$

For this model, the original DSC is used as a metric to evaluate the prediction results of the model, but a Soft Dice loss is used as a loss function for training. This metric was chosen because of its simplicity, ease of implementation and because it deals correctly with class imbalance, which is a big problem in medical image segmentation (since the prostate constitutes a small surface in the entire slice, a lot of black pixels will be present in the mask). Other losses have also been considered like the pixel accuracy and the Intersection over Union metric (IoU). The pixel accuracy fails to deal with class imbalance since it compares each pixel one by one with the ground truth [43] (so if the mask contains a lot of black pixels then the accuracy will most probably be high even though if the segmentation is completely failed). The IoU metric also called Jaccard's Index is simply based on the ration between the Intersection of pixels in both images, and the Union of the pixels. This metric is very useful in image segmentation and its concept is close to DSC's concept [43]. It will be used later on along other metrics, during the testing to evaluate the results of each model. Since the predicted masks need to have black and white pixels only, binary cross entropy was also envisaged. If dealt with class imbalance correctly, binary cross entropy could certainly be used as a loss function in this model. The ideal case would be to see in practice, which loss function would perform better for our data and this model. But due to time constraints and due to the fact that the goal of this master thesis is not to have the best model, but a working model, testing both loss functions and then choosing which is better would not only consume valuable time resources but also wouldn't focus on the dedicated goal of this research, considering that the Soft DSC leads to a working model.

Data augmentation as well as other parameters that affect the prediction of the model were also chosen during the building. These parameters will be presented and discussed in the next chapter as they consist an important part of training and testing the model.

2.3.3 Lesion Detection and Classification

In this section, the different steps that were followed to build, train and validate a model to detect prostate lesions on MRI scans and classify them as significant or not-significant, will be presented.

Pre-processing

All the MRI scans obtained from the **PROSTATEx** challenge came in the *DICOM* format, except the K-trans images that came in the *MHD* and *zraw* format [20]. As explained previously, this dataset provided a variety of MRI modalities compared to the **PROMISE12** dataset : T2-W (3 axes), PDW, DCE (K-trans), DWI (ADC map and

B-values images). The first steps of pre-processing were to choose which MRI images to keep as a basis for the study and which format was best for manipulating these images :

- As for the MRI modalities, axial T2-W, K-trans, B-val and ADC volumes were kept for this study. In terms of T2-W MRI, the axial plane cut is the favorite for a diagnosis of prostate diseases since it shows clearly the different zones and anatomy of the gland, that is why it was a logical choice to choose it over the other plane cuts. In addition to that, the other types of MRI are provided in that field of view also, so registering the image would be more logical to do if all images were taken from the same point of view. PDW images were not taken due to their unpopularity in diagnosing prostate cancer. In chapter 3, experiments will be held on the effect of removing or adding types of images to the learning
- Formatting the images to a specific type is necessary. First because K-trans come in a different format than the others and secondly, the *DICOM* format isn't suitable for manipulation since the 3D volumes come in 2D image slices, which means every time a case needs to be loaded, different 2D images need to be loaded instead of a single one containing the entire volume. That is why the decided format of manipulation is the *NIfTI* format which not only saves 3D volumes as a single file but also is easily loaded via different *Python* libraries which is the programming language used for this research. The *DICOM* images as well as the K-trans images were all converted to the *NIfTI* for simplicity and generalisation

The images were not standardised in terms of voxel spacing (spacing means the size of each voxel in the MRI volumes). In fact different subjects had different voxel spacing values which lead to a non-uniformity in the data. The third step in terms of pre-processing was to re-sample the voxel sizes to have a uniform size for each type of MRI.

To recap the previous step, after downloading the images from the online dataset, we filtered the MRI modalities that we needed for the study, we set a standard format for all images which is the *NIfTI* format and formatted all the images. Finally we re-sampled the the voxel spacing of the images so that each type could have one single voxel spacing.

The next steps involve downloading and pre-processing the masks to achieve a coherence between the images and the lesion masks. In fact, the **PROSTATEx** dataset doesn't provide corresponding masks to their images, where the lesions are visible. What they did provide is the coordinates of each lesion for each subject on the MRI images. The challenge here is that providing the location of a lesion in an MRI scan isn't equivalent to a mask of the lesion, where the mask covers the entire **volume** of the lesion over the different scans and not just indicates its location. A initial algorithm was built to try and solve this obstacle, it was based on an image segmentation method which is region growing. This region-based image segmentation method takes as input initial seed points in the segmented region of an image and tries to expand by selecting neighbouring pixels and analysing if these pixels are part of the region. This method would take as input the coordinates provided by the challenge and expand the region in the second and third dimensional spaces. The downside of this method is the accuracy. Not only MRIs are hard to segment to their non-uniformity, but prostate cancer isn't always a uniform shaped

mass, specially across the MRI volume. A second solution was envisaged to deal with this accuracy problem, which is using an already available masks dataset. This dataset of lesion masks for the **PROSTATEx** challenge was completed by a lesion-by-lesion quality check conducted at the Department of Advanced Biomedical Sciences of the University of Naples "Federico II" [44] [45]. In fact, two radiology residents and two experienced board-certified radiologists worked side by side on this dataset that they released in their research with a goal to provide publicly available lesion and prostate masks. After taking a close look at the masks provided in this dataset for each subject with an experienced urologist, the second solution was envisaged since it provided more accurate results than the first [46].

After acquiring the masks for the different lesions and subjects, here are the steps that were followed to pre-process these masks and make them ready for the training phase :

1. Re-sampled the voxel spacing of the masks to match the spacing of the MRI images
2. Register the masks as well as the other MRI modalities to the corresponding T2-W image for each patient. Registration is a full on process that aligns two similar images to have the same coordinate system. Here, since we have different MRI modalities of the same patient and the same organ from the same view, registering would help aligning these different MRI modalities and also align the mask to the corresponding image. This could highly benefit the training of the model since a region in one image will be easily identified in the other image (since they are aligned). In the case of the **PROSTATEx** dataset, the images were not aligned well enough, as well as the downloaded masks. T2-W axial images were chosen as the reference for the basis of the coordinates and registration was done between T2-W and the masks, as well as T2-W and the other MRI modalities (for each subject separately). Many algorithms for registration exist, here affine transformation was used in order to register the images
3. Similarly to the prostate segmentation data normalisation is applied (each MRI modality has its own normalisation values). Also here the normalisation was done using the training set
4. Prepared the data so that for each patient we have a volume containing the masks to the different lesions (or lesion) that the patient was diagnosed with. It was also essential to differentiate between the pixel representing the prostate, pixels representing non-significant lesions and pixels representing significant lesions
5. Cropped the masks and the training images to a size of 128x128 pixels around the prostate zone only. Cropping is an essential step here, since in the different prostate mp-MRIs, the prostate is visible along with its surroundings. Cropping the images around the prostate zone only (Region Of Interest) will help the model focus on the essential. It is important to add that this step is possible because of the previous model that segments the prostate in the MRI images :
 - (a) In terms of training, the masks provided in the training set contained already a segmentation of the prostate. This is really beneficial since segmenting the

prostate using the first model for the training images was not necessary, which would have of course propagated the error to the training set of the second model

- (b) In terms of the entire pipeline, we can clearly see here how both models connect and help each other out. The first model takes in the MRI, and segments the prostate. The second model uses the output of the first model to crop the prostatic region out of the MRI (ROI), which would help out the second model since it will be able to only focus on the prostate (ROI) and not its surroundings
 - (c) It is also important to justify the 128x128 cropping. As discussed previously, the prostate segmentation outputs a segmentation map of 256x256, and as discussed here, a cropping is necessary for the model to focus on the prostatic zone. After going through the different mp-MRIs of the different patients provided in the **PROSTATEx** dataset, it was concluded that a size of 128x128 was fitting correctly the prostate, without taking too much of its surroundings and without leaving out the external zones of the prostate. The number of parameters of the model was also taken into account in order to choose a standard cropping size of 128x128 since the goal was to start with a low amount of parameters and then augment if needed
6. Stacked the images up in a *Numpy* array as well as the masks, to make the training data ready for the model
 7. A last step was to do a one-hot-encoding to the masks. Since in each mask we could potentially have 3 types of pixels : one for the background, one for significant lesions and one for non-significant, it is important to create 3 masks from every mask each representing a category

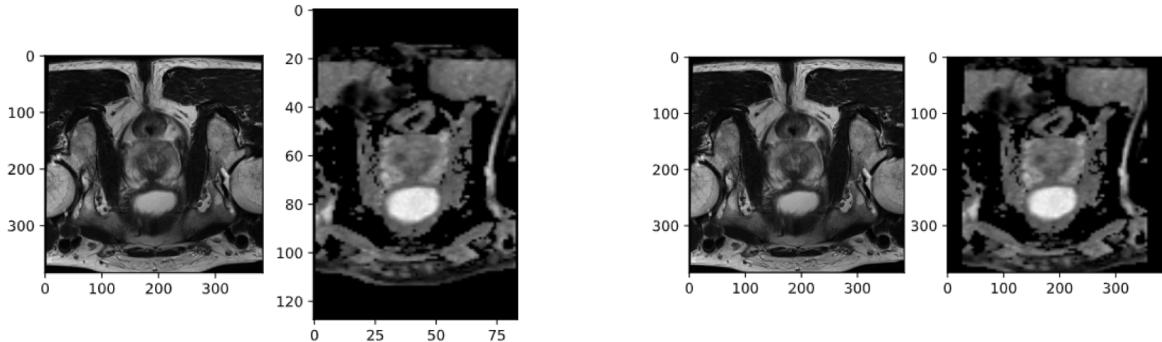


Figure 2.6: On the left a T2-W MRI and an ADC map not registered. On the right, the ADC map was registered to the T2-W MRI using affine transformation [20]

To sum up this pre-processing part on the **PROSTATEx** dataset, many steps were followed in order to clean the data, have the images in a universal and easy to use format, in a similar coordinate space, with the corresponding lesions and prostate masks. A last step involved cropping the images and the masks in order to focus on the ROI which is

the prostate gland. As you can see, pre-processing the images from the **PROSTATEx** dataset was not easy and took many weeks of work in order to have a format that could be manipulated easily. These steps were also essential in order to study the data and analyse it closely before using these images for training.

Network Structure

In terms of architecture, the U-Net model was used as a basis to the conception of the network structure for detecting and classifying lesions. The goal here is to take as input mp-MRI images of a prostate and complete a **semantic segmentation** of the lesions that appear to be significant (potentially cancerous) and the lesions that appear to be non-significant. The model should be able to first of all detect the lesions and at the same time classify each lesion in one of the two categories. As we have seen in the related work section of the previous chapter, many techniques exist to detect or segment lesions as well as many other exist to classify the detected lesions. Some of these techniques work independently (for example one network that segments the lesions and one CNN that takes as input a patch around the segmented lesion and classifies it as significant or not significant) and others merge the two tasks in one. In this research, the second option was adopted. The goal here is to build a network that would detect and classify the lesions in a single run, by adopting a semantic segmentation approach.

After building and testing the segmentation network presented in the previous section (the testing results will be detailed in the next chapter), it was concluded that this network accomplished accurate and satisfying results in terms of segmentation. Since both models should accomplish a similar task which is taking an MRI and detecting a specific region within it, it was decided to opt for a similar architecture in terms of the U-Net model. The major differences between the model built for prostate segmentation and the model built for lesion detection/classification are the input to the models, the output of the models and the size of the network. In fact, since the goal of this research is to build a full software pipeline, from segmenting the prostate to classifying the lesions, **the output of the previous model** was also used as an **input to the lesion detection/classification model** :

1. In terms of input, the model built for prostate segmentation only receives an axial T2-W-MRI of the prostate. For the second model, we first use the results of the prostate segmentation to zoom in on the ROI area which is the prostate gland in all the mp-MRIs (T2, ADC, K-trans, B-val). Since the prostatic lesions are generally located within the prostate, the idea here is to already indicate to the model the location of the prostate, to make the learning faster and generate better results
2. As previously explained, the mission of this second model is to detect the lesions and differentiate between significant and not significant lesions in a single run. For this reason, in terms of output, the model needs generate a mask with the lesions appearing as white pixels. But to be able to showcase the classification, cancerous lesions need to have a different pixel gray scale value compared to non-cancerous lesions. The illustration on figure 2.7 provides a better visualisation of a typical output of this model

3. Like said previously, the architecture of this model is based on the U-Net architecture. The same differences that were added to the prostate segmentation model that differ from the original U-Net model were also added here (padding, residual connections, variance scaling ...). But since the image input shape here is smaller (128x128 here compared to 256x256 for the prostate segmentation network) the depth of the network should be modified. As we have seen previously, in the original paper of the U-Net, the input image shape is divided by a factor of 20 compared to the image obtained at the bottom of the network. Here since our images have a shape of 128x128, we obtain a value of 6.4. In this network also, the number of channels in the first convolutional layer is 8, which indicates that the shape of the bottom image should be of a 4x4 value with a depth of 5, or an 8x8 shape with a depth of 4 since we want to get it close to 6.4. In this case, it was chosen to opt for the depth of 4 with a bottom image shape of 8x8. The reasons behind this choice are similar to previously : added to the fact that a 4x4 is relatively small, a depth of 5 could also yield good results and be correct, but since the goal of this master thesis isn't to test every possible parameter and obtain the best model, it was chosen to go for a depth of 4 and in case of none satisfying results, test the depth of 5
4. It is also important to mention that the activation function of the last layer of the network, which was a *Sigmoid* in the previous model, is now a *Softmax* for this model. The change was due to the number of classes in the output max, where in this model we have 3 classes (background, significant lesion, non-significant lesion) compared to previously where we only had to detect the prostate as white pixels (binary classification)

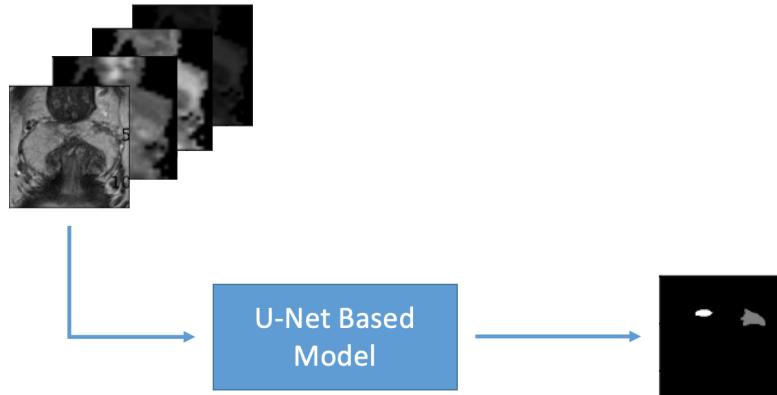


Figure 2.7: Representation of the different inputs fed to the model and the generated output [20]

In the next chapters of this thesis, this model will be presented as a model which achieves the detection as well as the classification of cancerous lesions. Even though these tasks may be done separately, it is important to mention here that the model built achieves both these tasks in a single run, by doing a semantic segmentation and segmenting

separately the lesions that are significant and the lesions that are not significant. It is totally possible to build two different models for these tasks (one that detects lesions and another that classifies them) but the approach developed in this thesis, which is a unique way of solving this problem, aims to accomplish both these tasks in a single run, using one single model that completes a semantic segmentation.

Training & Validation

As previously mentioned, the dataset consists of 344 subjects which divide in 204 training cases and 140 testing cases. Since the dataset was provided for an online challenge, the test set cannot be used here because the results were not made available. That is why the 204 cases were split as of 15% for testing and 85% for training and validation (10% of the 85%).

Each MRI volume made available in this dataset contains about 20 2D scans. But from these scans, very few contain a valid visualisation of prostatic lesions that could be used for training. Around 50% of each volume contains a prostate, and the lesions can be visualised in 60% of these 50% (if available). For training, validating and testing the model, only images containing a prostate were taken, and that is for the two reasons below :

- The images not containing a prostate, so not containing lesions are numerous. If we integrate them into the training set, we will have a big imbalance in the data since the number of images not containing lesions will be significantly higher than the number of images containing lesions. Instead of only taking images where we have lesions, we took images where we have a prostate
- Second, this is possible because the model built previously can indicate in which image we have a prostate, and in which we don't. Since the goal is to use both models in a single pipeline, what we are doing here is connecting the first model to the second by using the first model's output as an input to the second model , which will indicate to us in which slices of each volume we have a prostate, and in which we don't (since lesions can't be visible if we don't have a visible prostate)

In terms of numbers, the number of slices in the dataset that contain a prostate are at a number of 1885 slices (this number seems lower than the number given when presenting the datasets in chapter 2.2.2 but this is due to the fact that not all images containing **only** a prostate were taken, but 1 in 3 images were taken, and that is because the number of images containing only a prostate is much higher than the number of images containing lesions or prostate and lesions). Since 15% goes for testing (283), and 85% goes for training and validation (1602) we can directly conclude that data augmentation will probably be needed during training. This will be discussed later on in the next chapter.

Similarly to the prostate segmentation, DSC and Soft Dice Loss were used as a metric and as a loss function respectively. Indeed, since the task here is also a segmentation task, it was straightforward to use this kind of coefficient to evaluate the model, for the same reasons as mentioned in the previous section. Other metrics were also used to evaluate

the model and will be discussed in the next chapter.

Grad-CAM

When doctors analyse MRI scans, they don't only look at a particular mass or shape that could potentially be a cancerous region, but they also look at the surroundings and the entire region. Instead of only outputting a segmented shape with a score of how probable this lesion is cancerous, Grad-CAM technique was also used in this second model to provide additional localisation information of the lesions in order to offer better visualisation aspects to the radiologists and the doctors. Using this technique, the doctors won't just receive numerical value, but they will receive a representation of "how is the machine thinking? what is it analysing?" which would potentially help them form a more accurate and faster diagnosis.

Here are some examples of what Grad-CAM can show on prostate MRIs and located lesions. These examples were taken from predictions made by the model in charge of locating and classifying the lesions (second model). More details about the testing and the results of the model will be provided later on in the next chapter.

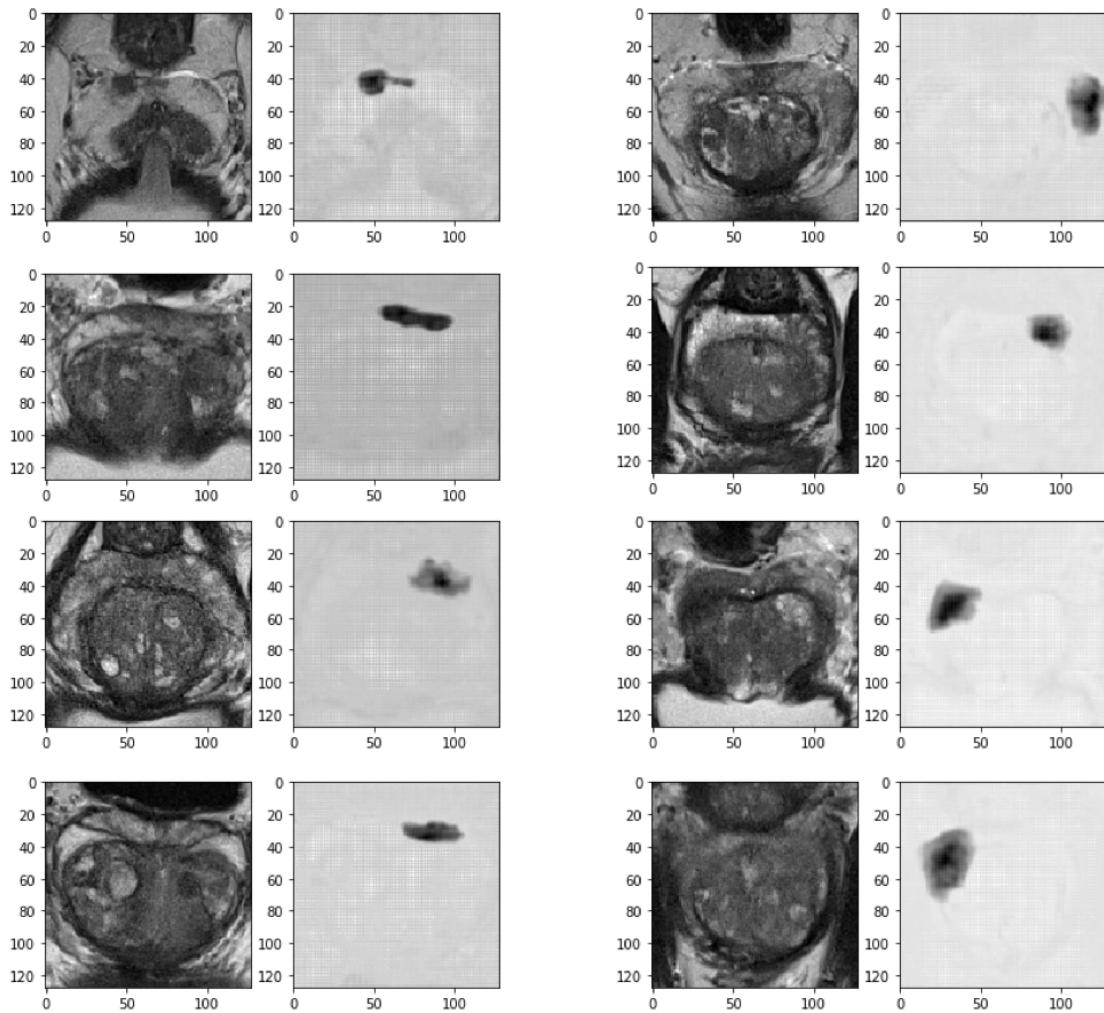


Figure 2.8: Grad-CAM results on the second model : detecting and classifying lesions within prostate MRI. In each image, on the left is represented the T2-W-MRI image and on the right the output of the grad-CAM map. The black regions on the Grad-CAM maps represent lesions in the prostate gland

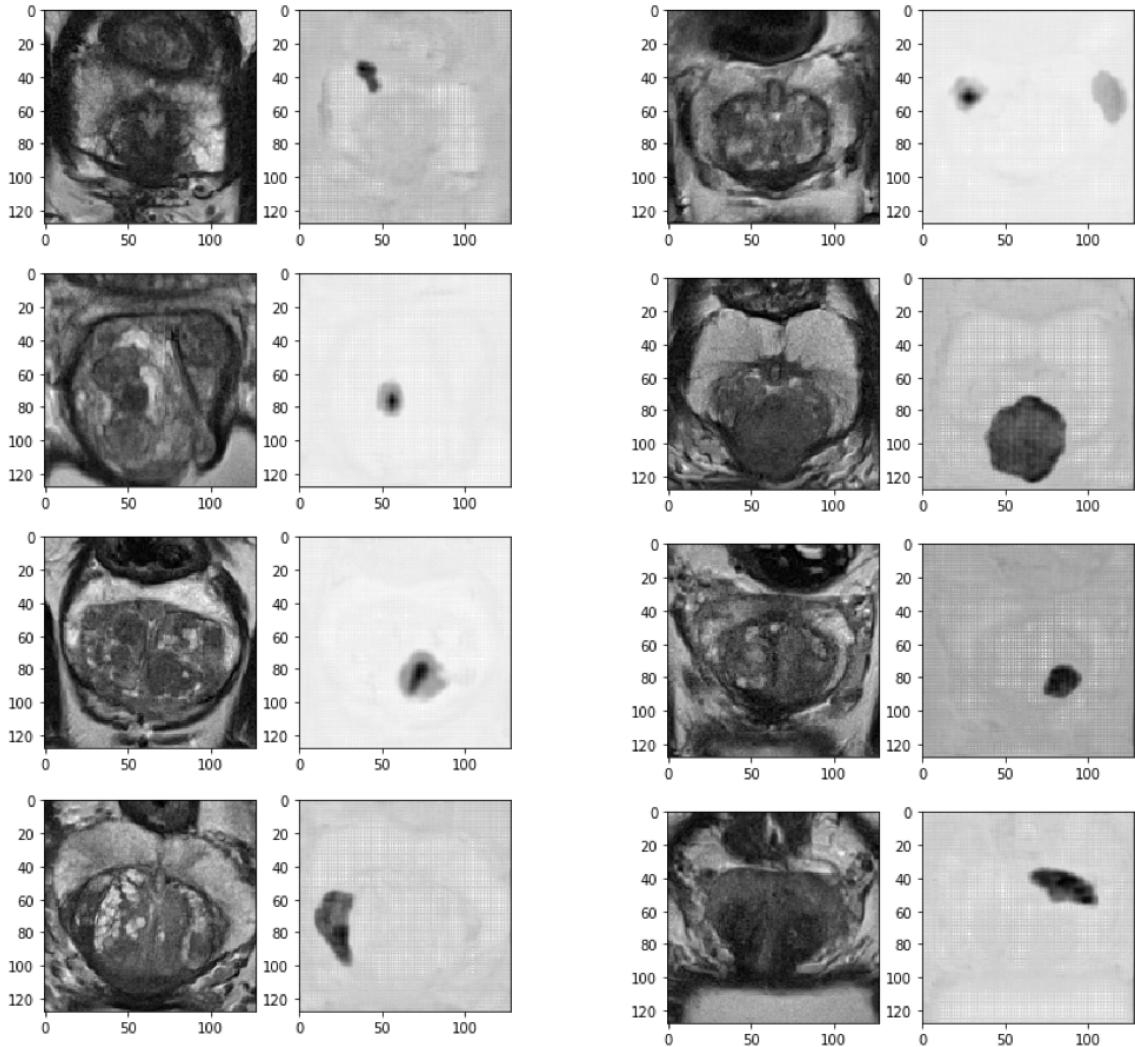


Figure 2.9: Grad-CAM results on the second model : detecting and classifying lesions within prostate MRI. In each image, on the left is represented the T2-W-MRI image and on the right the output of the grad-CAM map. The black regions on the Grad-CAM maps represent lesions in the prostate gland

2.3.4 Entire Pipeline

The goal of this Master Thesis research isn't only to segment a gland and its characteristics within MRI scans, but to also help radiologists and doctors with their diagnosis of prostate cancer. As said previously, Artificial Intelligence is doing a good job in the medical imaging field, but trusting a machine's diagnosis over a doctor's, is not foreseeable in the near future. In the mean time we cannot neglect the advantages that these new technologies are providing and the best way to do that is to form a synergy between human intelligence and machine intelligence. In the case of prostate cancer diagnosis, this research aims to accomplish this synergy by providing the following help : instead of building a software that would generate a complete diagnosis for detected prostate cancer in a patient's mp-MRI images, the software would generate insights that the doctor

would use to improve his diagnosis, and all of that via a non-invasive method. From these insights we can note :

- First, an insight on the size of the prostate, its shape and its localisation in the MRI. The first model segments the prostate on T2-W-MRI. This would help the doctors better visualise the prostate and even showcase some examples to their patients. It is also important to note that an increase in the prostate's size is a direct sign to PCa or a prostatic disease [4]. Segmenting a patient's prostate in a series of MRI taken over a period of time would allow the doctors to visualise the variation in the size and shape of the prostate
- Second, insights on lesion localisation and significance. A typical CAD software outputs its own diagnosis in terms of PCa : the lesion, its localisation and its significance. Doing that would totally replace the doctor's diagnosis and would lead to an AI taking over the work. Instead of doing that, in this research the software pipeline outputs a localisation map which indicates which regions in the MRI could probably be lesions as well as which lesions could probably be significant/not-significant. The difference here is, instead of outputting a numerical value, the software outputs what the machine is seeing, which would be considered by the doctor as an insight to his diagnosis instead of a competition to his work

Below is an overview of the entire pipeline, from what the radiologist or doctor using the software should input, to what he will get as an output :

1. As an input, the following mp-MRI scans should be provided : T2-W Axial, ADC map, K-trans and B-val images
2. The first step is the prostate segmentation. The model segments the prostate and outputs it as a mask covering the T2-W. The mask could also be applied to the other images
3. The second step is the lesions detection/classification. The model pre-processes the data, takes as input the mp-MRI images as well as the mask produced in the first step to zoom in on the ROI
4. As an output, the software delivers Grad-CAM localisation maps that indicate in each MRI where it believes a lesion is present and whether the detected lesion is significant or not

To conclude, we can compare the pipeline developed here to the models presented in the related works section 1.2. To first compare them, let's look at the output of these different methods. We have seen in the related works that most models complete only one specific task, either prostate segmentation, lesion detection or classification. Some models combine lesion detection and classification while others use a prostate segmentation model developed within another paper. Most of the papers prefer focusing on a single task and perfect it. The advantage of the approach developed in this research is that it combines these three tasks in one single paper. The goal here isn't to perfect a single task, but to try and combine these three tasks in a single pipeline. The second point of comparison is

the architecture. We saw in the related works that there exists papers that build models which achieve the three tasks of prostate segmentation, lesion detection and classification. What differs this research from the rest is the approach and the architecture :

- First of all the approach, since the focus of the research isn't to re-produce the doctor's diagnosis, but to help him by providing visual insights. In fact, by implementing Grad-CAM and outputting visual insights, this leads to a different approach from most of the papers on how to face the medical problematic situation
- In terms of architecture, the end goals are reached here by combining two U-Nets model, which are trained separately but connected in a way that the second model uses the first model to generate a ROI around the different mp-MRI images and focus only on the prostate by eliminating its surroundings
- To conclude, the uniqueness in this thesis doesn't come from building a model that will perfect the task of prostate segmentation or lesion detection/classification, but to build a software that will merge different ideas, challenges and datasets in order to build a pipeline that will achieve all of these three tasks that will help doctors in terms of PCa diagnosis

Chapter 3

Experiments and Results

Building a deep learning model doesn't stop at choosing an architecture. In fact, beyond the architecture many other parameters affect the performance of the model and its predictions. After introducing prostate cancer and the architectures that were built to try and help with the diagnosis of PCa from mp-MRI, in this section, the different experiments that were held will be presented. The goal of these experiments are not only to train the models and make them ready to predict, but to also see how these predictions can be improved via different techniques to achieve a better accuracy and avoid overfitting.

This third chapter is divided into three parts. The first two parts are dedicated to the different experiments that were held for each task of the pipeline, as well as the obtained results. The last section is an overview of the entire pipeline and the entire features that can be produced via this software. For the experiments that were held, the Dice Similarity Coefficient (DSC), the Intersection Over Union index (IOU), the Mean Relative Absolute Volume difference and Hausdorff Distance were used as metrics for studying the predictions of each model.

Here are some technical details useful for any reader that wishes to reproduce the experiments and obtained results :

- Programming Language : Python version 3.7.3 [47]
- Libraries used :
 - Tensorflow (v2) and Keras (2.3.1) (deep learning models) [48] [49]
 - Numpy (storing and manipulating the images) [50]
 - cv2 and SimpleITK (image processing) [51] [52]
- While building the code of the prostate segmentation as well as lesions detection and classification, I mainly inspired myself from two *github* repositories [39] [53]:
 - The first repository was built by Inom Mirzaev, from the Ohio State University [39]. The repository contains the source code of a project that aims to solve the **PROMISE12** challenge. I based myself on this repository in order to

imagine and build the architecture of my prostate segmentation code. I used different functions that were implemented in this repository in order to pre-process the MRI images from the dataset and make them ready for training. I also inspired myself by different implementation methods in this repository to build the code for the lesion detection and classification tasks. As I am a beginner in deep learning, I also used the code made available in this repository to help build the deep learning U-Net models

- A second repository I used was built by Alex Hamilton [53]. In this repository, Alex Hamilton tries to segment cancerous regions within mp-MRI images. Even though his methods and approach are really different from the method I was building, I used his code in order to pre-process the different MRI images that were made available within the **PROSTATEx** challenge. Due to the different types of images made available in the dataset, and the difficulty in processing and studying each type, pre-processing these mp-MRIS was necessary as explained earlier. This repository lead me on the right path on how to pre-process such data and make it ready for usage
- It is also important to mention that other sources were used as a basis of inspiration and help in order to build, train and validate the models as well as prepare the downloaded data : Keras documentation [49], TensorFlow documentation [48], Github Gist [54] and Stack Overflow [55]. A more detailed description on this matter can be found in the link below

Please note that the integrity of the source code developed during this work of research can be found here : <https://github.com/naimsassine/MasterThesisInfo2021>

3.1 Prostate Segmentation

As a reminder, the goal in this step if to take in a prostate T2-W-MRI and segment the prostate gland that is present in the image. The model that was built is based on the U-Net structure, with dropout and batch normalization layers. The training data consisted of 42 cases with 5 random cases chosen for validation, while the test set consisted of 8 cases. The soft Dice Loss was used a loss function for tracking the performance of the model (as discussed previously), Adam was used as an optimiser with a learning rate of 0.001 (choosing such a learning rate doesn't indicate that it is the best learning rate available for such a model. It was chosen to go for such a learning rate after different experiments were held to analyse the speed of the learning. Of course other learning rates could totally work better than this one, but as said, the goal of this master thesis isn't to search for each parameter in each model to build the best performing model, that is why it is useless to test every single value a learning rate could take, unless we really want to perfect the model down to each parameter, which is not the goal here).

The batch size used was 32 and validation was performed after each epoch. The best performing model in terms of validation was saved. As we have previously discussed, the number of images being insufficient for such a model (around 3 million trainable parameters with a training set of 1377 images), data augmentation was used in order to enlarge the size of the training set. Data augmentation consists of a set of techniques used to increase the set of data that is dealt with. These techniques introduce slight changes to the already existing data to diversify and enlarge the dataset, as well as develop synthetic data based on the original set. For example, if we take a 2D MRI scan of a prostate, we can rotate it, translate it, flip it, apply some transformations and from one image, create 10 more images. These transformations can also be applied randomly so that the model trains on a large variety of data, which will make it learn a more generalisable prediction.

To train this model using a data augmented training set, the *ImageDataGenerator* class from *Python's Keras* library was used to expand the size of the training set [56]. This class is very particular, because not only it generates augmented data from randomised transformation, but it also generates this data in real time, while the model is training. This makes the model more robust and saves a lot of memory (the size of the dataset is really non-negligible). It is also important to note that an elastic transform function [57] was used as a preprocessing function within the *ImageDataGenerator* class [56]. This function which is applied at each input help augment the data by applying a set of elastic distortions that will help augment the data and make the training more robust. Below are the training and validation score (DSC) and loss curves that were obtained when training the model on an augmented dataset.

Let us first look at the graphs generated from training the model. On the figure 3.1 we can see the curves of the dice score over the different epochs, in terms of training and validation. We can also see on the second plot, the variation of the dice loss over the epochs for training and validation too.

If we first take a look at the training curves, we can clearly see that the model is

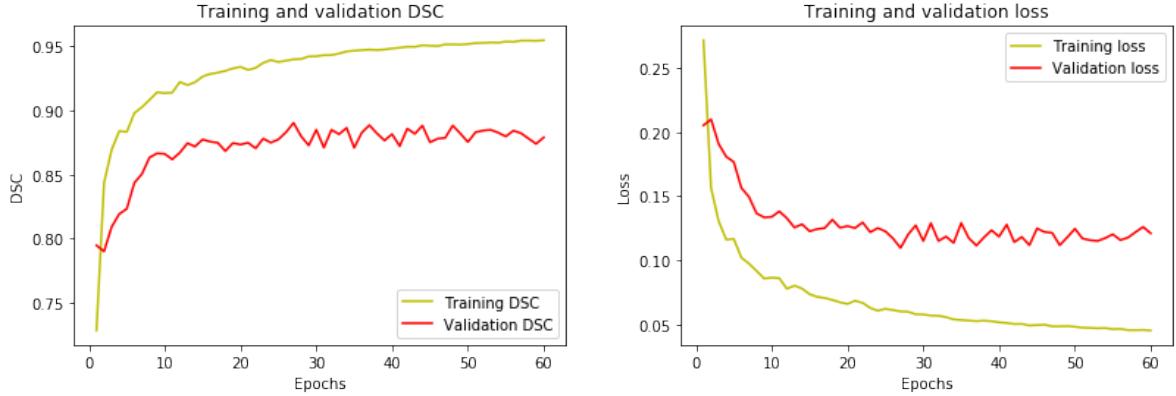


Figure 3.1: Training and validation dice similarity coefficients and DSC losses - Prostate Segmentation

achieving good training. It is learning how to localise the prostate and the different parameters behind the T2-W-MRI that will indicate the prostatic zone. For 60 epochs, the model keeps on learning and reaches a DSC of around 95% which is really good. By looking at the validation curve, we can also see that the validation scores are evolving in the same direction as the learning, for the first 20 epochs : the DSC is augmenting. After the first 20 epochs, the validation curves start to reach a certain plateau where the scores become somewhat constant. Even though the validation curves aren't as steady as the training curves (which could be due to many reasons such as a high learning rate, low amount of data or even some adjustments in the training augmentation data could help), we can clearly see that the validation curves reach a certain value before stopping their evolution, specially compared to the training curves which continue to evolve (around epoch 25 we can clearly see that the learning curve is still evolving, while the validation oscillates between a couple values, but stays in the entourage of 87% of DSC). This is probably a sign of overfitting : the model learned, but now it is starting to memorise the training data, which leads to an increase in the training accuracy but not in the validation accuracy. To be able to further study the aspect of the curves and in order to confirm that the model is overfitting after 60 epochs, more epochs have to be completed. Due to time constraints (1 epoch takes around 12mins, which leads to 12 hours for 60 epochs, 24 hours for 120), this couldn't be achieved, but since the goal here is to have a working model, let us look at the testing scores to see if the model is ready to be used :

- DSC : 0.87
- Mean IOU : 0.88
- Mean Relative Absolute Volume Difference : 12.59%
- Hausdorff Distance : 9.03 pixels

Before reflecting on the obtained results, let us dive into the chosen metrics and their significance :

- The DSC was introduced in the previous chapters. Please do keep in mind that the Soft Dice was used for training, but here as a metric we are using the typical DSC. As a reminder, the DSC is a score that measures the similarity between two masks or a set of masks, and it is scored from 0 to 1. A perfect similarity would achieve a score of 1
- The Intersection Over Union score was also introduced earlier. It is a metric that is known for measuring overlap between two masks [58]. Its formula is simple as it divides the areas of overlap of the two masks by the areas of union. Since this metric ranges from 0 to 1, a score of 1 would be ideal since it would mean that the area of intersection is equal to the area of union, which means that the compared masks are the same
- The Mean Relative Absolute Volume Difference, as its name indicates, is the difference between two mask volumes. What is done here is for each patient, we compare two volume of masks : the original and the predicted. In the **PROMISE12** dataset, each patient has his own volume of T2-W-MRI with the corresponding mask. So we get a single mask volume per patient. By using the built model we generate a prediction for each volume and we compare the predicted mask volume with the original mask volume. Once that is done for each patient, we compute a mean to get a single score. In fact that score gives us a view on how to model is doing on a patient per patient basis, instead of an overview on the entire prediction. Not to mention that is is really important for the model to be able to segment correctly the prostate in an entire volume and not just the slices separately, since MRIs are most often looked at by doctors in volumes. The smaller the percentage of difference is, the better, since the score of zero means that the volumes are identical
- The Hausdorff distance is the longest distance you can be forced to travel by an adversary who chooses a point in one of the two sets, from where you then must travel to the other set. In other words, it is the greatest of all the distances from a point in one set to the closest point in the other set [59]. In terms of image segmentation, this metric is really useful to measure how far away are situated the borders of the compared masks. This is useful since, compared to the DSC and the IOU, what is measured is not the similarity or the overlap, but how far in terms of distance are the two compared masks from each other. This comes in handy when the predicted mask doesn't have the exact same shape as the original. In this case the DSC and the IOU could be low (because the two shapes won't be totally overlapping) but the Hausdorff distance could indicate to us that the two shapes are near each other, and its just the borders that aren't correctly segmented. The Hausdorff Distance is expressed here in pixels and the smaller its value is, the better

The DSC score for the test set of this model, which has a goal of segmenting the prostate, reaches a value of 0.87. Compared to the training and validation curves, this shows that model has learned correctly, and generalised its learning to the testing set since a score of 0.87 yields a very good accuracy in terms of segmentation. The same conclusion applies to the mean IOU since we obtain a value of 0.88. The Mean Relative Absolute Volume Difference gives a total of 12.59% which is relatively low considering the number of pixels in each image (65 536 pixels). This shows that the model correctly

predicts prostate masks given a patient's volume. This also applies to the Hausdorff Distance as it gives a value of 9.03 pixels which is also a good value considering the size of the masks.

Here are some plotted MRI examples of the capabilities of the built model for prostate segmentation :

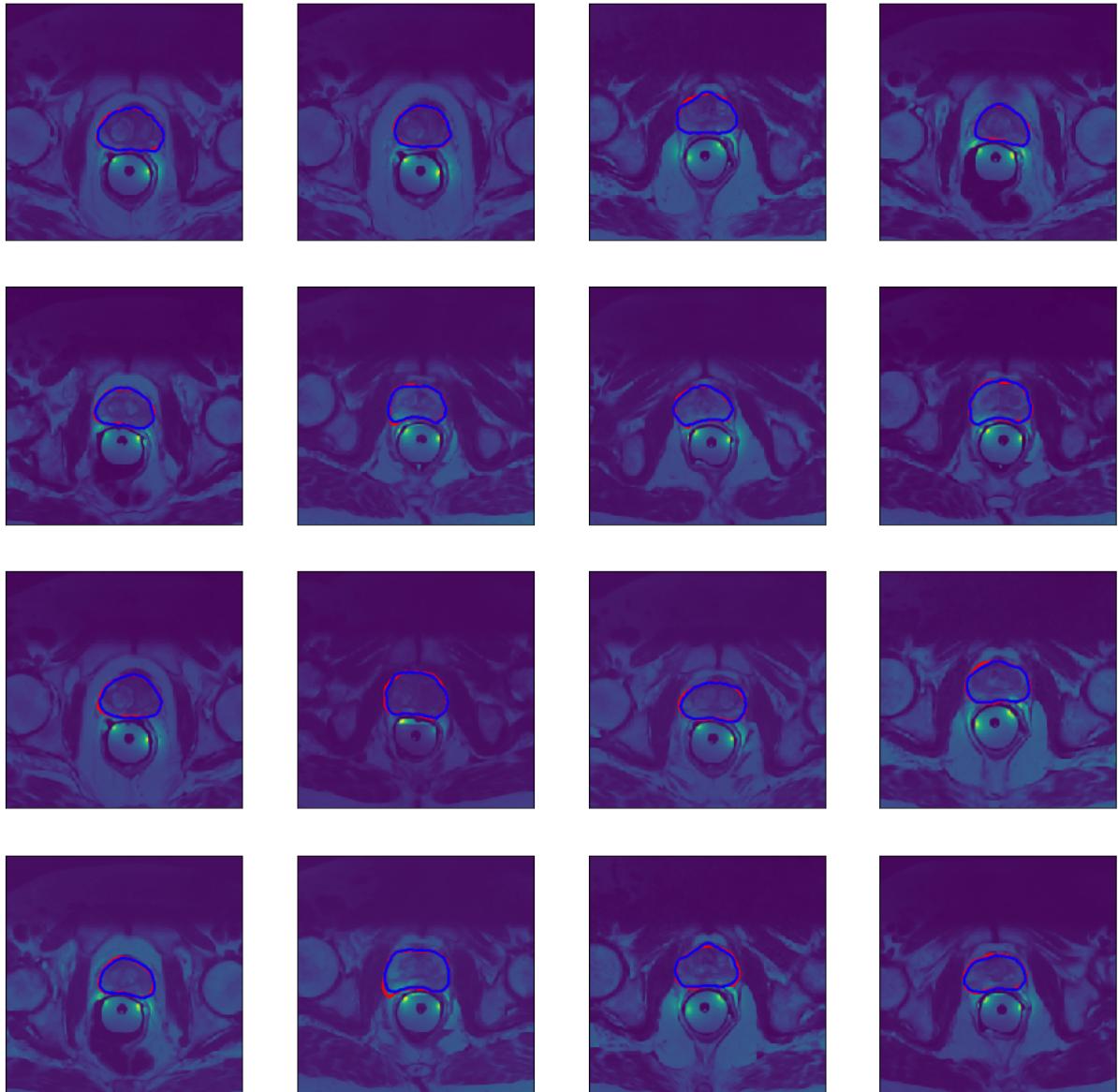


Figure 3.2: Segmenting the prostate on T2-W-MRI using the built deep learning model. In red are the original masks and in blue are the predicted ones

To sum up, we can say that the model built for prostate segmentation achieved its dedicated tasks. The initial goal was to build a model that would correctly segment the prostate, with a good score (the goal wasn't to get a perfect model, otherwise many more tests would have been necessary in order to optimise each parameter and generate better results). This goal was reached with a model that achieved an accuracy of 87% on the

testing set. We can conclude this first model of our pipeline by stating that we obtained a model that correctly segments the prostate in MRI images.

3.2 Lesion Detection and Classification

After successfully training and testing the first model that achieves prostate segmentation, this second section will be dedicated to training and testing the second model on the two tasks of lesion detection and classification in a single run (semantic segmentation). As previously explained, the goal here is to have a model that will take in a series of 2D MRI scans from different modalities (T2-W, K-trans, ADC and B-val) and output a mask where the lesions will be highlighted. Since two types of lesions are available in the training data (significant and non-significant) the model will also have to differentiate between both these types and highlight them differently, which is the classification task.

In terms of training parameters, these were similarly set as for the previous model : a learning rate of 0.001, soft Dice Loss as a loss function, batch size of 32 and validation performed after each epoch. The best performing model in terms of validation was also saved. As a reminder, 15% of the entire set of usable images was used for testing and 85% for training/validation (validation takes up 10% of these 85%), which leads to a total of 283 images for testing and 1602 for training/validation (1885 total images) . Since here also the number of images being insufficient for such a model (around 800000 trainable parameters with a training set of 1495 images), data augmentation was used in order to enlarge the size of the training set. The same data augmentation technique was used here compared to the previous model.

It is crucial to remind here that the **PROSTATEx** data set isn't well partitioned. In fact, the first few cases of this set contain non-significant lesions, compared to the last few cases that have a much higher rate of significant lesions. For that reason, a randomized shuffle was done to try and include an equal amount of significant and non-significant lesions in the training, validation and test sets. This would not only diversify the sets of data to improve the learning and evaluation, but would also make the model more robust and generalisable by eliminating an imbalance in the class weights. Please note that this randomisation prevents the use of the Mean Relative Absolute Volume Difference since this metric is based on a volume difference per patient, and here the slices of each patient will be mixed with each other for the different sets to deal with this problem of imbalance that the **PROSTATEx** dataset contains.

Let us first look at the graphs in figure 3.3 generated during the training of the model. These graphs are generated in the same fashion compared to the prostate segmentation graphs. In terms of the training curves, we can directly see that the model is learning. The training curve in terms of DSC is augmenting and not even reaching a plateau after 120 epochs. Even if the intensity of the curves starts getting smaller at 20 and 40 epochs, we can clearly see that the DSC score keeps augmenting. This usually is a good sign that the model is learning, but in order to truly evaluate what is happening, let us take a look at the validation curves. We can clearly see that the validation DSC score is also augmenting, but starts reaching a plateau at 100 epochs. Like previously said, this usually is a sign of overfitting since the training curve keeps growing but the validation starts stabilising around an interval of values. More epochs would be needed in order confirm that after 110 epochs the model starts overfitting, but due to time constraints,

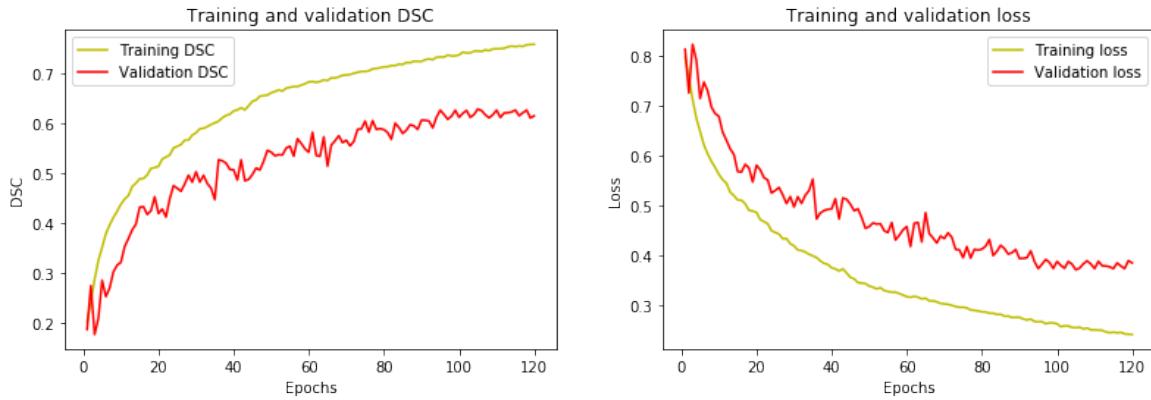


Figure 3.3: Training and validation dice similarity coefficients and DSC losses - Lesions Detection/Classification

this wasn't possible to accomplish. But since the goal here isn't to achieve the perfect model, let us analyse the validation curve and the testing results to see if this model is a valid and usable one.

Another non-negligible aspect of the validation curves is its instability. We can clearly see a lot of oscillations all along the validation curves, specially compared to the training one. Even though the aspect of the curve seems pretty clear (growing), the instability is present. This can be caused by different factors related to the lack of validation samples, the imbalance in the data, a high learning rate and could also potentially be fixed by optimizing data augmentation parameters. But since the aspect of the curve shows us that the model is learning, and the learning is generalising correctly to the validation set, we can directly look at the testing results to judge if this model is good enough or not.

- DSC : 0.61
- Mean IOU : 0.69
- Hausdorff Distance : 130.95 pixels
- Out of 193 Lesions, 136 are correctly detected (70%)
- Out of 139 Non-Significant Lesions, 92 are correctly detected (66%)
- Out of 54 Significant Lesions, 44 are correctly detected (81%)

The DSC score for the test set of this model reaches a value of 0.61. This accords with the validation curves obtained in the previous graph since the curve reaches a plateau at a value of 0.60. This first shows that the model correctly generalised to the testing set. A score of 0.61 (over 1.0) could definitely be improved by optimising and tuning different parameters. But the lack of training images and the imbalance in the data makes this training difficult. The Mean IOU confirms this analysis with a score of 0.69. Compared to the prostate segmentation, the Hausdorff Distance hits a higher score of 130.95 (compared to the previous 9.03 of the prostate segmentation). Even though this

score may seem high, it is important to evaluate it taking into consideration the image size. The Hausdorff Distance is computed per pixel, so a distance of 1 represents a single pixel. The masks have a size of 128x128, the size of the lesions differ a lot from an image to another and from a patient to another, but they usually take up 25% of the image, which can be estimated to a size of 32x32 :

- On a scale of 128x128, 130.95 pixels represent 0,7%
- On a scale of 32x32, 130.95 pixels represent 12%

We can conclude that on both scales, the distance between the predicted lesion and the correct lesion is relatively small and that the Hausdorff Distance yields a good score. This indicates that the lesions are close to each other, and that the prediction is correctly indicating the location of the lesions.

If we look at the number of correctly predicted lesions, we can see that the model achieves good result by detecting around 70% of the lesions (by detecting it is meant here, knowing that this MRI contains a lesion). It also clearly well identifies between significant and none-significant lesions.

From these different results, we can conclude that :

- The model is learning and generalising correctly
- The model is able to recognize the different lesions
- The model is able to differentiate between non-significant and significant lesions
- The model is able to detect these lesions (relatively good Hausdorff distance and percentage of correctly detected lesions), but segmenting them accurately seems like a more difficult task to accomplish (relatively low DSC and IOU)

Let us look at some one screen examples to verify these conclusions.

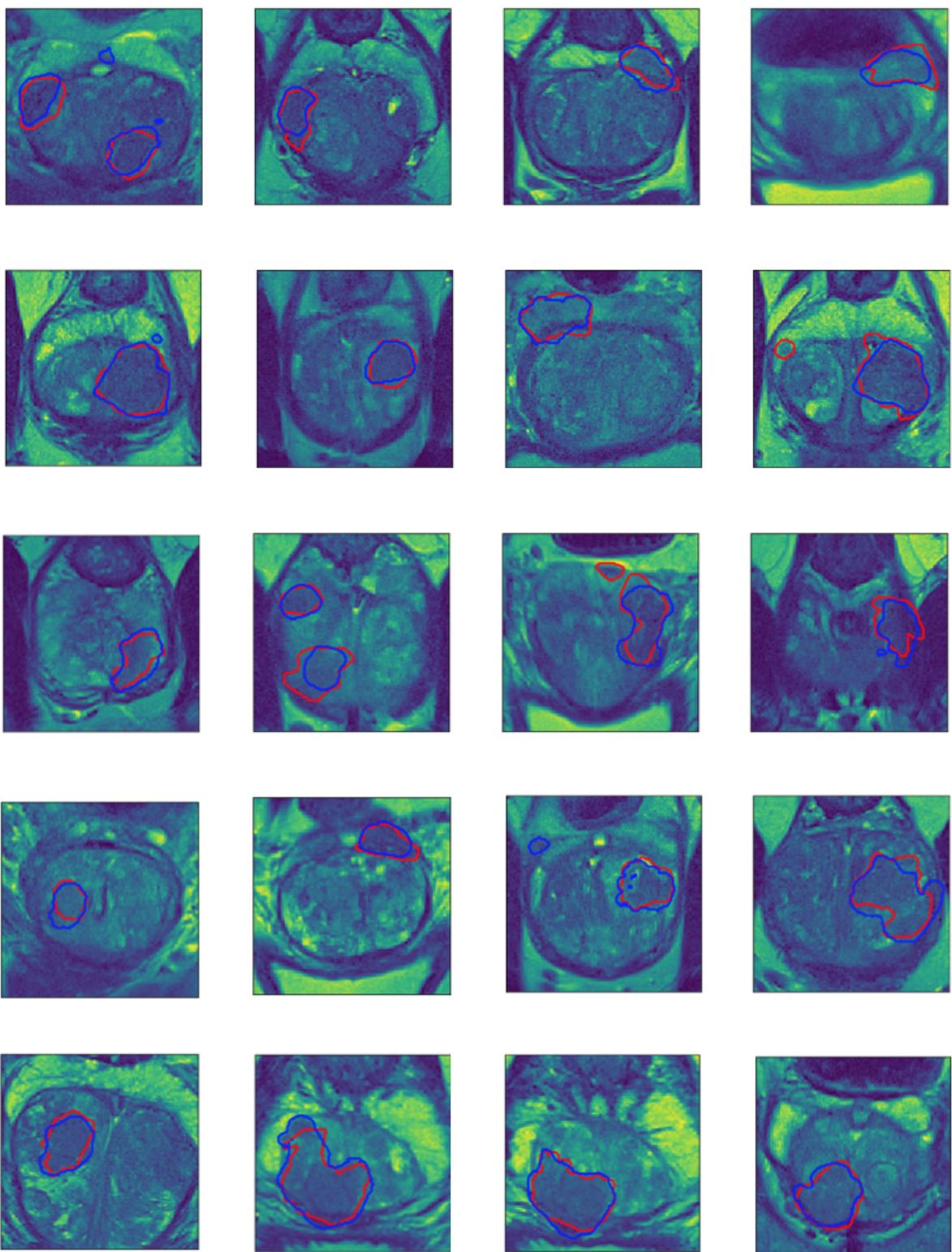


Figure 3.4: Semantic segmentation (detection and classification) of lesions within the prostate gland on T2-W-MRI images using the built deep learning model. Only non-significant lesions are represented here. In red are the original masks and in blue the predicted ones

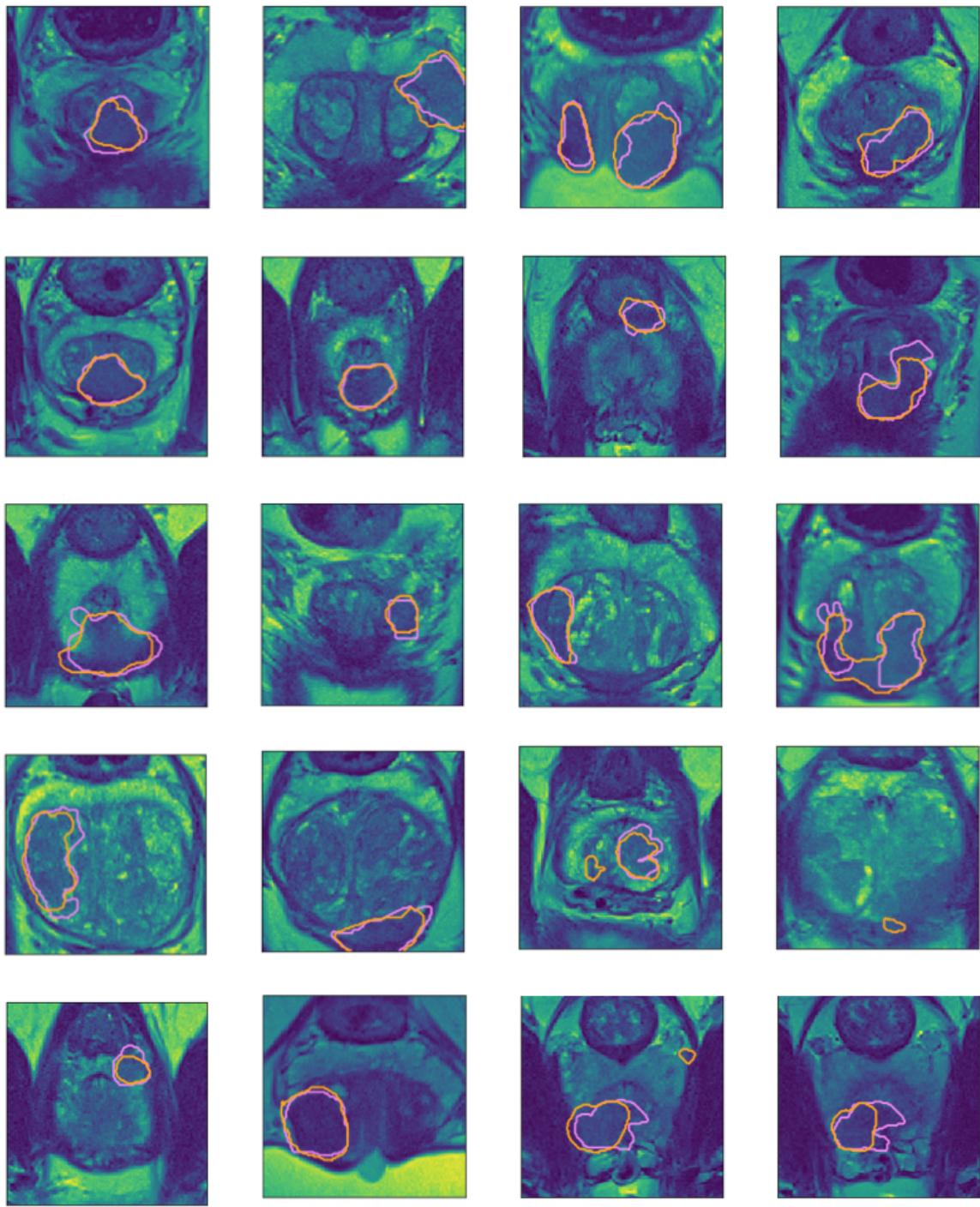


Figure 3.5: Semantic segmentation (detection and classification) of lesions within the prostate gland on T2-W-MRI images using the built deep learning model. Only significant lesions are represented here. In violet are the original masks and in orange the predicted ones

By looking at the above examples, we can see that the conclusions drawn previously are confirmed by the produced segmentation masks. First, we can clearly see that the

model is learning and able to differentiate between significant lesions and non-significant lesions. Then, we can visualise the results and metrics obtained previously : the model is able to learn where the lesions are, it is able to predict their location as well as their significance, but in terms of segmentation, the model is having a hard time segmenting every pixel correctly, specially the edges of the lesions. This justifies the good Hausdorff distance, but the relatively low DSC, as the model locates the lesions but doesn't segment the edges in an accurate way.

The initial goal for this model was to obtain a network that would indicate the location of lesions within prostate MRIs and predict if each lesion is significant or not. The goal wasn't to segment every pixel of every lesion at a high accuracy. Since the global objective is to build a software for doctors, indicating the location of the lesion and its significant are the most important tasks that need to be achieved since once indicated, the radiologist can by hand segment it if needed. It is important to remind the reader here that the goal wasn't to obtain a perfect model that would yield a 100% accuracy in segmenting the lesions, but to have a model that would locate and classify lesions in a patient's MRI scans, and based on the previous results, that goal is clearly achieved by this model.

3.3 Entire Pipeline

After building, training and testing these two models, an overview of the entire software pipeline will be presented in this section : from the radiologist's input to the software's output.

As a reminder, the goal of this research is to build a software that would provide useful insights for radiologists and doctors to help with the diagnosis of prostate cancer, via non-invasive methods :

1. A radiologist completes a prostate MRI scan on one of his patients. The images taken during this process are of different types : T2-W, ADC, K-trans, B-val ...
2. The radiologists takes the images that were produced from the MRI scan and uploads them to the software
3. The software firstly takes in the T2-W-MRI volume of this particular patient, and uses the previously built and trained network to automatically segment the prostate in these images. The segmentation is then printed out to the doctor
4. The doctor can now visualise the different MRI scans with a mask on top of the prostatic gland in the images. If the patient already went through an MRI scan a while back, the doctor can benefit from the masks to see the increase/decrease in the prostate's size, which is a key element in the diagnosis of prostate diseases
5. After predicting the segmentation of the prostate, the software zooms in on this specific region of the image, and crops all the images in accordance with the prostatic surface to be able to focus on the region of interest, which is here the prostate gland
6. The software then takes the four following MRI modalities : T2-W, ADC, K-trans and B-val, applies the necessary registering (spacing, registering, cropping ..) to then input them into the second deep learning model
7. This second model takes in this set of images and predicts, if available, the location of the lesions in these image modalities. The model not only detects and locates these lesions but also indicates if the lesion is clinically significant (cancerous) or not
8. As an output to this network, the software outputs the regions detected as lesions and their significance, as well as a heat map via Grad-CAM, which will give more insight to the doctor on what the network sees, and where
9. The insights provided can be used by the doctor as a push start in the diagnosis of PCa or other types of prostate diseases

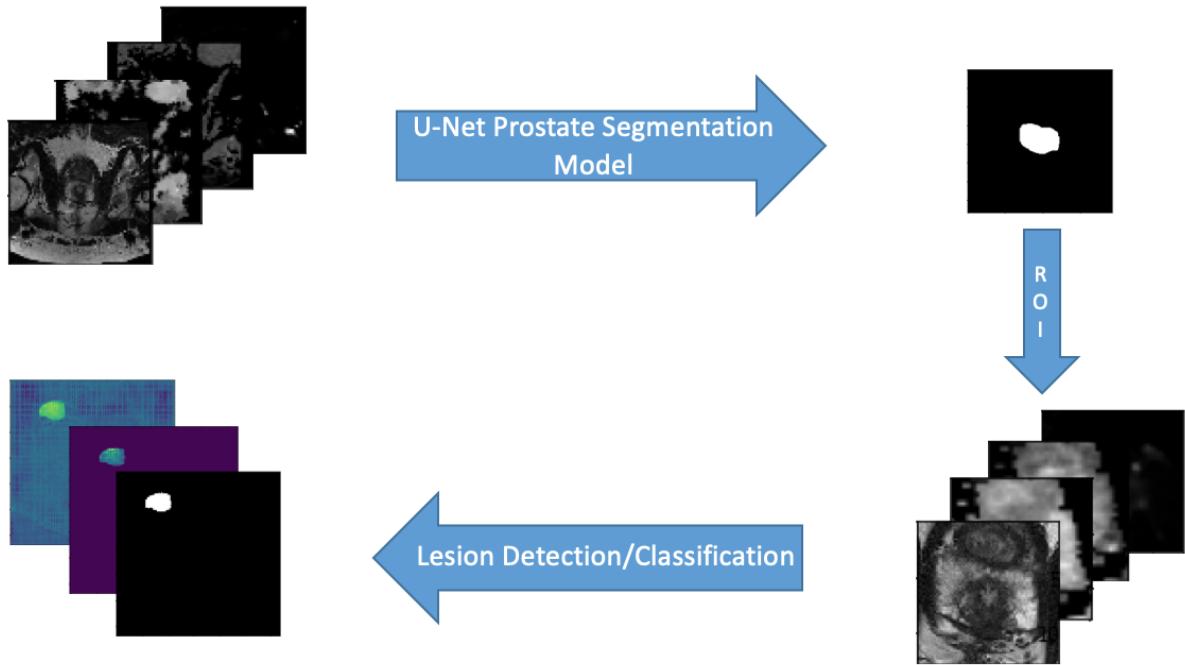


Figure 3.6: The entire pipeline of the software

We have seen above that both models yield good metrics and results in terms of the goals that were set to be achieved in the beginning of this thesis. But, although these goals were evaluated with the previous metrics, the best way to evaluate a software built to be used by doctors is to test out the software in a hospital or clinic and see if doctors would be able to generate a better diagnosis with the software. Even though this seems like the ideal method to see if the initial goal of the software is attained, testing it with doctors in an establishment is beyond the scope of a Master's Thesis.

Chapter 4

Conclusion

Prostate cancer (PCa) is one of the most common types of cancer diagnosed worldwide. Often showing no symptoms until a later stage, PCa is not easily detectable. Invasive methods are the most common techniques used to diagnose this type of disease. Non-invasive methods, like MRI screening, can also be used to diagnose PCa but often lead to erroneous results due to the different interpretations of abnormalities within these images. The goal of this research was to develop a software based on deep learning, that studies the different prostate MRI modalities of a patient and outputs insights that would help the doctors in achieving a better diagnosis, via a non-invasive method.

The software focuses on three tasks : prostate segmentation, lesion detection and lesion classification. In order to complete these tasks, two deep learning models were conceived, trained, validated and tested. In chapter 2, the building block of these models, which is the U-Net architecture, was presented as well as the datasets that were used to train and test the two models. The second section of chapter 2 served as an overview of the architecture of both these models. All the pre-processing steps were also discussed in order to clarify how the data was downloaded and prepared.

The first model takes in a T2-W-MRI of a prostate and segments the gland within the image. The second model takes in 4 different types of MRI modalities (T2-W, ADC, K-trans and B-val), uses the results of the first model to crop the images into a region of interest, then in a single run detects the lesions and classifies them as cancerous or not cancerous. The different details related to the training and validation of both these models were presented in chapter 3. Different experiments were held to train the models, evaluate them and fine tune different parameters to reach the initial goals, of building a working CAD pipeline, usable by doctors, to help them in their PCa diagnosis.

In terms of prostate segmentation, a Dice Similarity Coefficient (DSC) of 87% was achieved on the test set. The second model completed the one run detection/classification task with a DSC of 61% on the test set.

As of future works, many improvements are still awaiting in this field. In fact, a multitude of experiments can still be performed in order to improve the two models. Testing if the detection of lesions gets easier when removing an MRI modality, applying

contrast enhancement filters, or even using the prostate mask created by the first model as an additional input to the second are all different experiments that could drastically improve the robustness of both models. A human touch integration can also be a way of building a new synergy between human and machine intelligence (for example the model takes also as input the doctor's opinion).

To conclude, a full software pipeline was built throughout this research that aims to help doctors diagnose prostate cancer via a non-invasive method. The different insights that this software outputs are related to prostate segmentation, which helps studying the difference in volume over time of a patient's prostate; lesion detection which helps the doctors identify lesions that are not always perceivable by the human eye; and lesion classification which helps radiologists with the task of assigning a label to each lesion based only on the MRI modalities.

Bibliography

- [1] Urology Care Foundation. *Prostate Cancer: Symptoms, Diagnosis & Treatment*. URL: <https://www.urologyhealth.org/urologic-conditions/prostate-cancer> (visited on 05/04/2021).
- [2] Hyuna Sung, Jacques Ferlay, et al. “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. In: *A Cancer Journal for Clinicians - Wiley Online Library* 71 (3 2021). URL: <https://acsjournals.onlinelibrary.wiley.com/doi/full/10.3322/caac.21660> (visited on 05/04/2021).
- [3] Canadian Cancer Society. *The prostate*. URL: <https://www.cancer.ca/en/cancer-information/cancer-type/prostate/prostate-cancer/the-prostate/?region=on> (visited on 05/12/2021).
- [4] Guillaume Lemaitre, Robert Marti, et al. “Computer-Aided Diagnosis for Prostate Cancer using Multi-Parametric Magnetic Resonance Imaging”. In: 39th International Conference of the IEEE Engineering in Medicine and Biology Society. Universite de Bourgogne; Universitat de Girona: IEEE Comput. Soc, 2016, p. 230. ISBN: hal-01516245. URL: <https://hal.inria.fr/hal-01516245> (visited on 05/19/2021).
- [5] American College of Radiology. *PI-RADS*. URL: <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/PI-RADS> (visited on 05/05/2021).
- [6] Ahmad Chaddad, Tamim Niazi, et al. “Predicting Gleason Score of Prostate Cancer Patients Using Radiomic Analysis”. In: *frontiers in Oncology* (2018). DOI: <https://doi.org/10.3389/fonc.2018.00630>. (Visited on 05/05/2021).
- [7] *PROSTATEx-2 Challenge*. URL: <https://www.aapm.org/GrandChallenge/PROSTATEx-2/> (visited on 05/05/2021).
- [8] Michelle D. Bardis, Roozbeh Houshyar, et al. “Applications of Artificial Intelligence to Prostate Multiparametric MRI (mpMRI): Current and Emerging Trends”. In: *Cancers* 12.5 (2020), p. 1204. DOI: <https://doi.org/10.3390/cancers12051204>. (Visited on 05/12/2021).
- [9] Leonardo Rundo, Carmelo Militello, et al. “Automated Prostate Gland Segmentation Based on an Unsupervised Fuzzy C-Means Clustering Technique Using Multispectral T1w and T2w MR Imaging”. In: *Information* 8.49 (2017). DOI: <https://doi.org/10.3390/info8020049>. (Visited on 05/05/2021).

- [10] Yi Zhu, Rong We, et al. “Fully automatic segmentation on prostate MR images based on cascaded fully convolution network”. In: *JMRI* 49.4 (2019), pp. 1149–1156. DOI: <https://doi.org/10.1002/jmri.26337>. (Visited on 05/05/2021).
- [11] *U-Net: Convolutional Networks for Biomedical Image Segmentation*. URL: <https://arxiv.org/abs/1505.04597> (visited on 05/28/2021).
- [12] Fausto Milletari and Nassir Navab. *V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation*. URL: <https://arxiv.org/abs/1606.04797> (visited on 05/05/2021).
- [13] Ke Yan et al. “A propagation-DNN: Deep combination learning of multi-level features for MR prostate segmentation”. In: *Computer Methods and Programs in Biomedicine* 170 (Mar. 2019), pp. 11–21. ISSN: 01692607. DOI: [10.1016/j.cmpb.2018.12.031](https://doi.org/10.1016/j.cmpb.2018.12.031). URL: <https://linkinghub.elsevier.com/retrieve/pii/S016926071831099X> (visited on 05/05/2021).
- [14] Nathan S. Lay, Yohannes Tsehay, et al. “Detection of prostate cancer in multiparametric MRI using random forest with instance weighting”. In: *Journal of Medical Imaging* 4.2 (2017), p. 024506. DOI: <https://doi.org/10.1117/1.JMI.4.2.024506>. (Visited on 05/05/2021).
- [15] Yohan Sumathipala, Nathan Lay, et al. “Prostate cancer detection from multi-institution multiparametric MRIs using deep convolutional neural networks”. In: *Journal of Medical Imaging* 5.4 (2018), p. 044507. DOI: <https://doi.org/10.1117/1.JMI.5.4.044507>. (Visited on 05/05/2021).
- [16] Saining Xie, Zhiowen Tu, et al. “Holistically-Nested Edge Detection”. In: *International Journal of Computer Vision* 125 (2017), pp. 3–18. DOI: <https://doi.org/10.1007/s11263-017-1004-z>. (Visited on 05/12/2021).
- [17] Yohannes K. Tsehay et al. “Convolutional neural network based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonance images”. In: SPIE Medical Imaging. Ed. by Samuel G. Armato and Nicholas A. Petrick. Orlando, Florida, United States, Mar. 3, 2017, p. 1013405. DOI: [10.1117/12.2254423](https://doi.org/10.1117/12.2254423). URL: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2254423> (visited on 05/05/2021).
- [18] Helen Xu, John S. H. Baxter, et al. “Prostate Cancer Detection using Residual Networks”. In: *International Journal of Computer Assisted Radiology and Surgery* volume 14 (2019), 1647–1650. DOI: <https://doi.org/10.1007/s11548-019-01967-5>. (Visited on 05/05/2021).
- [19] Saifeng Liu et al. “Prostate cancer diagnosis using deep learning with 3D multiparametric MRI”. In: SPIE Medical Imaging. Ed. by Samuel G. Armato and Nicholas A. Petrick. Orlando, Florida, United States, Mar. 3, 2017, p. 1013428. DOI: [10.1117/12.2277121](https://doi.org/10.1117/12.2277121). URL: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2277121> (visited on 05/05/2021).
- [20] *SPIE-AAPM-NCI PROSTATEx Challenges*. URL: <https://wiki.cancerimagingarchive.net/display/Public/SPIE+-AAPM+-NCI+PROSTATEx+Challenges> (visited on 05/14/2021).

- [21] Atilla P. Kiraly et al. “Deep Convolutional Encoder-Decoders for Prostate Cancer Detection and Classification”. In: *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*. Ed. by Maxime Descoteaux et al. Vol. 10435. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 489–497. ISBN: 978-3-319-66178-0 978-3-319-66179-7. DOI: 10.1007/978-3-319-66179-7_56. URL: http://link.springer.com/10.1007/978-3-319-66179-7_56 (visited on 05/05/2021).
- [22] Zhiyu Liu et al. “A Two-Stage Approach for Automated Prostate Lesion Detection and Classification with Mask R-CNN and Weakly Supervised Deep Neural Network”. In: *Artificial Intelligence in Radiation Therapy*. Ed. by Dan Nguyen, Lei Xing, and Steve Jiang. Vol. 11850. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019, pp. 43–51. ISBN: 978-3-030-32485-8 978-3-030-32486-5. DOI: 10.1007/978-3-030-32486-5_6. URL: http://link.springer.com/10.1007/978-3-030-32486-5_6 (visited on 05/05/2021).
- [23] Ross Girshick et al. “Rich feature hierarchies for accurate object detection and semantic segmentation”. In: *arXiv:1311.2524 [cs]* (Oct. 22, 2014). arXiv: 1311 . 2524. URL: <http://arxiv.org/abs/1311.2524> (visited on 05/13/2021).
- [24] Kaiming He et al. “Mask R-CNN”. In: *arXiv:1703.06870 [cs]* (Jan. 24, 2018). arXiv: 1703.06870. URL: <http://arxiv.org/abs/1703.06870> (visited on 05/13/2021).
- [25] Shaoqing Ren et al. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”. In: *arXiv:1506.01497 [cs]* (Jan. 6, 2016). arXiv: 1506 . 01497. URL: <http://arxiv.org/abs/1506.01497> (visited on 05/13/2021).
- [26] Yixuan Yuan et al. “Prostate cancer classification with multiparametric MRI transfer learning model”. In: *Medical Physics* 46.2 (Feb. 2019), pp. 756–765. ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.13367. URL: <https://onlinelibrary.wiley.com/doi/10.1002/mp.13367> (visited on 05/05/2021).
- [27] *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*. URL: <https://arxiv.org/abs/1610.02391> (visited on 05/23/2021).
- [28] Joachim Feger. *Prostate MRI protocol | Radiology Reference Article*. Radiopaedia. URL: <https://radiopaedia.org/articles/prostate-mri-protocol?lang=us> (visited on 07/21/2021).
- [29] *MRI Basics*. URL: <https://case.edu/med/neurology/NR/MRI%20Basics.htm> (visited on 07/21/2021).
- [30] *MRI Basics*. URL: <https://case.edu/med/neurology/NR/MRI%20Basics.htm> (visited on 05/28/2021).
- [31] MRI Master. *Proton Density (PD) weighted MRI sequence physics and image appearance*. URL: <https://mrимaster.com/characterise%20image%20pd.html> (visited on 07/21/2021).
- [32] *Home - PROMISE12 - Grand Challenge*. URL: <https://promise12.grand-challenge.org/> (visited on 05/14/2021).

- [33] Geert Litjens and Robert Toth. “Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge”. In: *Medical Image Analysis* 18.2 (2014), pp. 359–373. DOI: <https://doi.org/10.1016/j.media.2013.12.002>. (Visited on 05/28/2021).
- [34] Harshall Lamba. *Understanding Semantic Segmentation with UNET*. Medium. Feb. 17, 2019. URL: <https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47> (visited on 07/21/2021).
- [35] Jeremy Zhang. *UNet Line by Line Explanation*. Medium. Oct. 18, 2019. URL: <https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5> (visited on 07/21/2021).
- [36] Naoki Shibuya. *Up-sampling with Transposed Convolution*. URL: <https://naokishibuya.medium.com/up-sampling-with-transposed-convolution-9ae4f2df52d0> (visited on 05/21/2021).
- [37] Olaf Ronneberger, Fischer Philipp, and Brox Thomas. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. Vol. 9351. Series Title: Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24573-7 978-3-319-24574-4. DOI: 10.1007/978-3-319-24574-4_28. URL: http://link.springer.com/10.1007/978-3-319-24574-4_28 (visited on 07/21/2021).
- [38] Sumit Gupta. *High Accuracy & Faster Deep Learning with High Resolution Images & Large Models*. Medium. Aug. 20, 2019. URL: <https://sumitgup.medium.com/deep-learning-with-high-resolution-images-large-models-44bfd90482a8> (visited on 08/09/2021).
- [39] Inom Mirzaev. *PROMISE12 Challenge - Automated Segmentation of Prostate Structures from MR Images*. original-date: 2017-12-12T18:19:43Z. Aug. 3, 2021. URL: https://github.com/mirzaevinom/promise12_segmentation (visited on 08/09/2021).
- [40] *SimpleITK: CurvatureFlowImageFilter*. URL: https://simpleitk.org/doxygen/latest/html/classitk_1_1simple_1_1CurvatureFlowImageFilter.html#details (visited on 08/08/2021).
- [41] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: (), p. 8.
- [42] *What are “residual connections” in RNNs?* Cross Validated. URL: <https://stats.stackexchange.com/questions/321054/what-are-residual-connections-in-rnns> (visited on 07/22/2021).
- [43] Ekin Tiu. *Metrics to Evaluate your Semantic Segmentation Model*. Medium. Oct. 3, 2020. URL: <https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2> (visited on 07/21/2021).
- [44] Renato Cuocolo and Renato Comelli. “Deep Learning Whole-Gland and Zonal Prostate Segmentation on a Public MRI Dataset”. In: *Journal of Magnetic Resonance Imaging* 54 (2021), pp. 452–459. DOI: <https://doi.org/10.1002/jmri.27585>. (Visited on 05/22/2021).

- [45] Renato Cuocolo, Arnaldo Stanzione, et al. “Quality control and whole-gland, zonal and lesion annotations for the PROSTATEx challenge public dataset”. In: *European Journal of Radiology* 138 (2021), p. 109647. DOI: <https://doi.org/10.1016/j.ejrad.2021.109647>. (Visited on 05/22/2021).
- [46] Renato Cuocolo. *Lesion and prostate masks for the PROSTATEx training dataset, after a lesion-by-lesion quality check*. GitHub. URL: https://github.com/rcuocolo/PROSTATEx_masks (visited on 07/21/2021).
- [47] *Welcome to Python.org*. Python.org. URL: <https://www.python.org/> (visited on 08/09/2021).
- [48] *TensorFlow*. TensorFlow. URL: <https://www.tensorflow.org/> (visited on 08/09/2021).
- [49] *Keras: the Python deep learning API*. URL: <https://keras.io/> (visited on 08/09/2021).
- [50] *NumPy*. URL: <https://numpy.org/> (visited on 08/09/2021).
- [51] *OpenCV*. OpenCV. URL: <https://opencv.org/> (visited on 08/09/2021).
- [52] *SimpleITK*. URL: <https://simpleitk.org/> (visited on 08/09/2021).
- [53] *alexhamiltonRN - Github repository*. GitHub. URL: <https://github.com/alexhamiltonRN> (visited on 04/26/2021).
- [54] *Github*. GitHub. URL: <https://github.com> (visited on 08/09/2021).
- [55] *Stack Overflow - Where Developers Learn, Share, & Build Careers*. Stack Overflow. URL: <https://stackoverflow.com/> (visited on 08/09/2021).
- [56] *ImageDataGenerator*. TensorFlow. URL: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator (visited on 07/28/2021).
- [57] P.Y. Simard, D. Steinkraus, and J.C. Platt. “Best practices for convolutional neural networks applied to visual document analysis”. In: *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings*. Seventh International Conference on Document Analysis and Recognition. Vol. 1. Edinburgh, UK: IEEE Comput. Soc, 2003, pp. 958–963. ISBN: 978-0-7695-1960-9. DOI: [10.1109/ICDAR.2003.1227801](https://doi.org/10.1109/ICDAR.2003.1227801). URL: <http://ieeexplore.ieee.org/document/1227801/> (visited on 07/28/2021).
- [58] Oleksii Sheremet. *Intersection over union (IoU) calculation for evaluating an image segmentation model*. Medium. Sept. 7, 2020. URL: <https://towardsdatascience.com/intersection-over-union-iou-calculation-for-evaluating-an-image-segmentation-model-8b22e2e84686> (visited on 07/29/2021).
- [59] *Hausdorff distance*. URL: https://en.wikipedia.org/wiki/Hausdorff_distance (visited on 07/29/2021).